Current Biology

Insights into the Evolution of Multicellularity from the Sea Lettuce Genome

--Manuscript Draft--

Manuscript Number:	CURRENT-BIOLOGY-D-18-00475R3
Full Title:	Insights into the Evolution of Multicellularity from the Sea Lettuce Genome
Article Type:	Research Article
Corresponding Author:	Olivier De Clerck Ghent University Ghent, BELGIUM
First Author:	Olivier De Clerck
Order of Authors:	Olivier De Clerck
	Shu-Min Kao
	Kenny A. Bogaert
	Jonas Blomme
	Fatima Foflonker
	Michiel Kwantes
	Emmelien Vancaester
	Lisa Vanderstraeten
	Eylem Aydogdu
	Jens Boesger
	Gianmaria Califano
	Benedicte Charrier
	Rachel Clewes
	Andrea Del Cortona
	Sofie D'Hondt
	Noe Fernandez-Pozo
	Claire M. Gachon
	Marc Hanikenne
	Linda Lattermann
	Frederik Leliaert
	Xiaojie Liu
	Christine A. Maggs
	Zoe A. Popper
	John A. Raven
	Michiel Van Bel
	Per K.I. Wilhelmsson
	Debashish Bhattacharya
	Juliet C. Coates
	Stefan A. Rensing
	Dominique Van Der Straeten

	Assaf Vardi
	Lieven Sterck
	Klaas Vandepoele
	Yves Van de Peer
	Thomas Wichard
	John H. Bothwell
Abstract:	We report here the 98.5 Mbp haploid genome (12,924 protein coding genes) of Ulva mutabilis, a ubiquitous and iconic representative of the Ulvophyceae or green seaweeds. Ulva's rapid and abundant growth makes it a key contributor to coastal biogeochemical cycles; its role in marine sulfur cycles is particularly important because it produces high levels of dimethylsulfoniopropionate (DMSP), the main precursor of volatile dimethyl sulfide (DMS). Rapid growth makes Ulva attractive biomass feedstock, but also increasingly a driver of nuisance 'green tides'. Additionally, ulvophytes are key to understanding evolution of multicellularity in the green lineage. Furthermore, morphogenesis is dependent on bacterial signals, making it an important species to study cross-kingdom communication. Our sequenced genome informs these aspects of ulvophyte cell biology, physiology and ecology. Gene family expansions associated with multicellularity are distinct from those of freshwater algae. Candidate genes are present for the transport and metabolism of DMSP, including some that arose following horizontal gene transfer from chromalveolates. The Ulva genome offers, therefore, new opportunities to understand coastal and marine ecosystems, and the fundamental evolution of the green lineage.

Insights into the Evolution of Multicellularity from the Sea Lettuce Genome

Authors: Olivier De Clerck[1]*, Shu-Min Kao[2,3], Kenny A. Bogaert[1], Jonas Blomme[1,2], Fatima Foflonker[4], Michiel Kwantes[5], Emmelien Vancaester[2,3], Lisa Vanderstraeten[6], Eylem Aydogdu[2,3], Jens Boesger[5], Gianmaria Califano[5], Benedicte Charrier[7], Rachel Clewes[8], Andrea Del Cortona[1,2,3], Sofie D'Hondt[1], Noe Fernandez-Pozo[9], Claire M. Gachon[10], Marc Hanikenne[11], Linda Lattermann[5], Frederik Leliaert[1,12], Xiaojie Liu[1], Christine A. Maggs[13], Zoë A. Popper[14], John A. Raven[15,16], Michiel Van Bel[2,3], Per K.I. Wilhelmsson[9], Debashish Bhattacharya[4], Juliet C. Coates[8], Stefan A. Rensing[9], Dominique Van Der Straeten[6], Assaf Vardi[17], Lieven Sterck[2,3], Klaas Vandepoele[2,3,19], Yves Van de Peer[2,3,18,19], Thomas Wichard[5], John H. Bothwell[20]

[1] Biology Department, Ghent University, 9000 Ghent, Belgium.
[2] Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium.
[3] VIB Center for Plant Systems Biology, Technologiepark 927, 9052 Ghent, Belgium.
[4] Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ 08901, USA.
[5] Institute for Inorganic and Analytical Chemistry, Jena School for Microbial Communication, Friedrich Schiller University Jena, Lessingstr. 8, 07743 Jena, Germany.
[6] Laboratory of Functional Plant Biology, Department of Biology, Ghent University, K.L. Ledeganckstr. 35, 9000 Ghent, Belgium.
[7] Morphogenesis of Macroalgae, UMR8227, CNRS-UPMC, Station Biologique, Roscoff 29680, France.
[8] School of Biosciences, University of Birmingham, Edgbaston, B15 2TT, UK.
[9] Faculty of Biology, University of Marburg, Karl-von-Frisch-Str. 8, 35043 Marburg, Germany.
[10] Scottish Association for Marine Science, Scottish Marine Institute, Oban, PA37 1QA, UK.
[11] InBioS-Phytosystems, University of Liège, 4000 Liège, Belgium.
[12] Botanic Garden Meise, Nieuwelaan 38, 1860 Meise, Belgium.
[13] School of Biological Sciences and Queen's University Marine Laboratory Portaferry, Queen's University Belfast, Northern Ireland, BT7 1NN, UK.
[14] Botany and Plant Science, and, Ryan Institute for Environmental, Marine and Energy Research, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland.
[15] Division of Plant Sciences, University of Dundee at the James Hutton Institute, Dundee, DD2 5DA, UK.
[16] School of Biological Sciences, University of Western Australia (M048), 35 Stirling Highway, WA 6009, Australia.
[17] Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel.
[18] Department of Genetics, Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa.
[19] Bioinformatics Institute Ghent, Ghent University, Technologiepark 927, 9052 Ghent, Belgium
[20] School of Biological and Biomedical Sciences and Durham Energy Institute, Durham University, Durham, DH1 3LE, UK.
* Correspondence: olivier.declerck@ugent.be

Summary

We report here the 98.5 Mbp haploid genome (12,924 protein coding genes) of *Ulva mutabilis*, a ubiquitous and iconic representative of the Ulvophyceae or green seaweeds. *Ulva*'s rapid and abundant growth makes it a key contributor to coastal biogeochemical cycles; its role in marine sulfur cycles is particularly important because it produces high levels of dimethylsulfoniopropionate (DMSP), the main precursor of volatile dimethyl sulfide (DMS). Rapid growth makes *Ulva* attractive biomass feedstock, but also increasingly a driver of nuisance 'green tides'. Ulvophytes are key to understanding the evolution of multicellularity in the green lineage, and *Ulva* morphogenesis is dependent on bacterial signals, making it an important species to study cross-kingdom communication. Our sequenced genome informs these aspects of ulvophyte cell biology, physiology and ecology. Gene family expansions associated with multicellularity are distinct from those of freshwater algae. Candidate genes are present for the transport and metabolism of DMSP, including some that arose following horizontal gene transfer from chromalveolates. The *Ulva* genome offers, therefore, new opportunities to understand coastal and marine ecosystems, and the fundamental evolution of the green lineage.

Key words: green seaweeds, multicellularity, phytohormones, DMSP, DMS, *Ulva*

Introduction

Transitions from microscopic, unicellular life forms to complex multicellular organisms are relatively rare events but have occurred in all major eukaryotes lineages, including animals, fungi, and plants [1, 2]. Algae, having acquired complex multicellularity several times independently, provide unique insights into the underlying mechanisms that facilitate such transitions. In the green lineage, there have been several independent transitions to multicellularity. Within Streptophyta, the origin of complex land plants from a green algal ancestor was preceded by a series of morphological, cytological, and physiological innovations that began long before the colonization of land [3]. Within the Chlorophyta, the transition from uni- to multicellularity has occurred in several clades and has been studied particularly extensively in the volvocine lineage [4]. Comparative genomic analyses between unicellular (*Chlamydomonas*), colonial (*Gonium*, *Tetrabaena*) and multicellular species (*Volvox*) revealed protein-coding regions to be very similar, with the notable exceptions of the expansion of gene families involved in extracellular matrix (ECM) formation and cell-cycle regulation [5-7].

The Ulvophyceae or green seaweeds represent an independent acquisition of a macroscopic plant-like vegetative body known as a thallus. Ulvophyceae display an astounding morphological and cytological diversity [8], which includes unicells, filaments, sheet-like thalli, as well as giant-celled coenocytic or siphonous seaweeds [9]. *Ulva*, or sea lettuce, is by far the best-known representative of the Ulvophyceae, and is well established as a model organism for studying morphogenesis in green seaweeds [10]. The *Ulva* thallus is relatively simple, with small uninucleate cells and a limited number of cell types. *Ulva* exists in the wild in two forms: either as flattened blades that are two cells thick or as tubes one cell thick (Figure 1A). Both forms co-occur in most clades as well as within single species. These morphologies, however, are only established in the presence of appropriate bacterial communities [11]. In axenic culture conditions, *Ulva* grows as a loose aggregate of cells with malformed cell walls (Figure 1B). Only when exposed to certain bacterial strains (e.g., *Roseovarius* and *Maribacter*), or grown in conditioned medium, is complete morphogenesis observed [11]. Thallusin, a chemical cue inducing morphogenesis, has been characterized for the related genus *Monostroma* [12] and additional substances that induce cell division (*Roseovarius*-factor) and cell differentiation (*Maribacter*-factor) have been partially purified from *Ulva*-associated bacteria [11]. The value of *Ulva* as a model organism for green seaweed morphogenesis is enhanced by its tractability for genetic analyses. Under laboratory conditions, individuals readily complete the life cycle, which involves an alternation of morphologically identical haploid gametophytes and diploid

sporophytes [10]. Furthermore, a stable polyethylene glycol (PEG)-based genetic transformation system is available [13].

The *Ulva* genus is also of great ecological significance. *Ulva* is widely distributed along tropical and temperate coasts, and several species penetrate into freshwater streams and lakes. Under high nutrient conditions, *Ulva* can give rise to spectacular blooms, known as "green tides", often covering several hundreds of kilometres of coastal waters. Beached algae may amount to a million tons of biomass and smother entire coastlines and negatively impact tourism and local economies [14]. Although not toxic, green tides have led to fatalities due to the hydrogen sulphide that is formed when the dead biomass decays. Despite the harmful consequences of *Ulva*'s rapid growth rate, beneficial aspects include exploitation of its biomass; e.g., for biofuel production, protein for animal feed and the removal of excess nutrients in integrated multitrophic aquaculture systems [15].

The genomic resources for the Ulvophyceae that could shed light on their independent transition to multicellularity and the evolution of the cyto-morphological diversity are limited to a transcriptomic study of *Ulva linza* [16] and a description of the mating type locus of *U. partita* [17]. In addition, Ranjan and colleagues [18] studied the distribution of transcripts in the thallus of the siphonal green seaweed *Caulerpa*. Here, we present the first whole genome sequence of an *Ulva* species, *U. mutabilis* Føyn (Figure 1). The species is phylogenetically closely related to *U. compressa* Linnaeus (Figure S1), a widespread species known to form nuisance blooms. The sequenced strain was isolated from southern Portugal [19] and has been successfully maintained in culture. Several important aspects of its biology, including cell cycle, cytology, life cycle transition, induction of spore and gamete formation and bacterial-controlled morphogenesis, have been studied in detail over the past 60 years [10]. We intend our *Ulva* genome, in combination with the availability of a genetic toolkit and developmental and life cycle mutants, to spur the new developmental research in green seaweeds that will be critical if we are to harness their aquaculture potential, while mitigating the threat from blooms. Our analyses of the *Ulva* genome reveal features that underpin the development of a multicellular thallus, and expansion of gene families linked to the perception of photoperiodic signals and abiotic stress, which are key factors for survival in intertidal habitats. Furthermore, we unveil key genes involved in the biosynthesis of dimethylsulfoniopropionate (DMSP) and dimethyl sulfide (DMS), important signaling molecules with a critical role in the global sulfur cycle.

Results and Discussion

Genome sequencing and gene family evolution

The genome size of *U. mutabilis* was estimated by flow cytometry and k-mer spectral analysis to be around 100 Mbp. In total, 6.9 Gbp of PacBio long reads were assembled into 318 scaffolds (98.5 Mbp), covering 98.5 % of the estimated nuclear haploid genome (Figure 2, Table S1). To increase the accuracy of the genome sequence at single-base resolution, the scaffolds were polished using PacBio, and Illumina paired-end reads. We predicted 12,924 protein-coding genes, of which 91.8 % were supported by RNA-Seq data. Analyses of genome completeness indicated that the genome assembly captures at least 92% of the eukaryotic BUSCO dataset. Analyses of pico-PLAZA Core Gene Families resulted in a completeness score of 0.968 of the protein-coding genes (Table S2). Annotation of repetitive elements resulted in 35% of the genome being masked. Among the identified repeats, 74% were classified as known or reported repeat families, with long-terminal repeats (LTRs) and long interspersed elements (LINEs) being predominant, representing 15.3 Mbp and 9.3 Mbp, respectively (Table S3).

The *Ulva* genome size is intermediate between sequenced genomes in the Chlorophyceae and Trebouxiophyceae (Figure 2). The number of predicted genes and gene families are markedly lower compared to most Chlorophyceae, including the volvocine algae (*Chlamydomonas*, *Gonium, Tetrabaena* and *Volvox*), but higher than the Trebouxiophyceae and prasinophytes (Figures 2, 3). The relative gene family sizes, however, are roughly equal between *Ulva* and volvocine algae when corrected for total genome size (Figure S2). A phylogenetic tree inferred from a concatenated alignment of 58 nuclear protein-coding genes (totaling 42,401 amino acids) supports a sister-group relationship of *Ulva* with the Chlorophyceae in the crown chlorophytes (Figure 3). This topology corroborates earlier phylogenetic hypotheses based on multigene organelle datasets (reviewed in [8]). The divergence of *Ulva* and the *Chlamydomonas* – *Gonium* – *Volvox* clade (Chlorophyceae, Volvocales) from their common ancestor coincided with substantial gain and loss of gene families in both lineages (Figure 3). Gain and loss, however, do not seem to be correlated with multicellularity.

Evolution of multicellularity

Ulva develops from gametes or zoospores into a multicellular thallus consisting of three main cell types (rhizoid, stem and blade cells). After a first division, the basal cell gives rise to a rhizoidal cell and the apical

cell [20]. Subsequent morphogenesis is then governed by a progressive change in cell cycles. The growth rate of the basal cells decreases after a few cell cycles, so that the holdfast remains small, while the division of blade cells continues and becomes synchronised to the prevailing light:dark cycle [20]. The shape of the multicellular thallus in *Ulva* is therefore largely driven by how cell size and division are controlled and many morphological mutants in *Ulva mutabilis*, including the slender mutant used in this study (see below), appear to have arisen from underlying changes in cell cycle regulation [20].

The evolution of a complex thallus morphology is often associated with expansions in gene families that are involved in cell signaling, transcriptional regulation and cell adhesion [1]. The *Ulva* genome encodes 251 proteins involved in transcriptional regulation, a comparatively low number for a green alga, which is also reflected in a low fraction of such proteins encoded by the genome (1.94% when compared to the average of 2.66% in green algae). *Ulva* lacks ten families of transcription factors (TF) and two families of transcriptional regulators (TR) that are present in other green algae (Data S1). Furthermore, the existing transcription-associated protein families are, on average, smaller than those in other green algae (Figure 4A).

Among the most remarkable gene families that have been lost are genes of the retinoblastoma/E2F pathway and associated D-type cyclins. Comparative genomic studies of volvocine algae have revealed that the co-option of the retinoblastoma cell cycle pathway is a key step towards multicellularity in this group of green algae [5-7]. Apart from implying that evolution toward multicellularity progressed along different trajectories in *Ulva* and the volvocine algae, the absence of D-type cyclins, retinoblastoma (RB) and E2F signifies that entry into the cell cycle, and the G1-S transition, are independent on these genes, as is the case in yeast. As no homologs of Cln 2/3, SBF and Whi5, which mediate G1-S transition in yeast [21], are found in *Ulva*, we hypothesise that either a functionally analogous set of genes or an entirely different mechanism regulates *Ulva* S-phase entry. Interestingly, retinoblastoma and E2F homologs were found in the transcriptome of the siphonous ulvophyte *Caulerpa* [18], so at present it remains unclear how widely the retinoblastoma/E2F pathway is conserved within the Ulvophyceae. Other aspects of the cell cycle are more in line with other green algae (Data S1, Figure S4), be it that the single CDKA homolog (UM001_0289) contains a modified cyclin-binding motif, PSTALRE, instead of the evolutionarily conserved PSTAIRE motif. While variations in this motif are not uncommon in eukaryotes, *Ulva* is the first member of the green lineage to have such a variation reported.

Contrary to the expectation for a multicellular organism [22], few TF families are expanded in *Ulva* (Figure 4A). A notable exception are CONSTANS-LIKE transcription factors (CO-like) of which *Ulva* has five genes,

whereas all other sequenced algae encode between zero and two (Figure 4A, B). These CONSTANS-LIKE transcription factors are characterised by one or two (Group II or III) Zinc finger B-boxes and a CCT protein domain. Both protein domains are involved in protein-protein interactions, and the CCT domain mediates DNA-binding in a complex with HEME ACTIVATOR PROTEIN (HAP)-type transcription factors in *Arabidopsis* [23]. *Ulva* CO-like proteins form a single clade within other algal lineages (Figure 4B). In addition to the five CO-like transcription factors, functionally related proteins either containing a) only a B-Box domain similar to Group V B-Box zinc fingers [23] or b) only a CCT domain and belonging to the CCT MOTIF FAMILY [24], are also expanded in *Ulva*. B-box Zn fingers and CMF proteins in angiosperms have been implicated in developmental processes such as photoperiodic flowering [25], regulation of circadian rhythms [26], and abiotic stress responses [27]. The control of light and photoperiod signalling is conserved in the green algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri* [28, 29]. Moreover, the CO-like TFs are one of the families potentially involved in the establishment of complex multicellularity in green algae and land plants [30]. Genome-wide mapping of *Ulva* CO-like genes and functionally related genes indicates that the majority (60%) originated through tandem duplication in *Ulva* (Figure 4C). Although the functions of these proteins will need to be confirmed experimentally, the *CO*-like and *CMF* genes in *Ulva* could be involved in the integration of a multitude of environmental signals in a highly dynamic intertidal environment, of a kind to which the other sequenced green algae are not regularly subjected.

A total of 441 protein kinases were identified in the *Ulva* genome, representing ca. 3-4% of all protein coding genes. The largest subfamily of *Ulva* kinases has similarity to PKnB kinase (Data S1), a “eukaryotic-like” serine/threonine kinase originally discovered in bacteria. Around 20 of the PKnB kinases possess a transmembrane (TM) domain and an extracellular/adhesion domain: either Kringle (IPR000001), FAS1 (fasciclin-like; IPR000782), or Pectin lyase fold (IPR012334, IPR011050), and so represent good candidates for *Ulva* receptor-kinases, with potential roles in environmental sensing and/or developmental signalling. By acting on *Ulva* cell wall components, for example, the pectin lyase fold domains may contribute to desiccation resistance [31] and to the growth and development of a multicellular thallus [32]. We note that although the three aforementioned extracellular domains are present in many green algae, including the green seaweed *Caulerpa* (Figure 5), it is unusual to see them linked to a kinase, and this coupling is only seen in a few other eukaryotic species. The Kringle-kinase domain combination, for instance, was initially discovered in animal receptor tyrosine kinase-like orphan receptors (RORs), which use Wnt signalling proteins as ligands and function in multicellular development, neuronal outgrowth, cell migration and polarity [33]. Our analysis additionally finds this domain combination in *Ulva* and the unicellular prasinophyte green algae (*Ostreococcus*). The Fasciclin-kinase combination is unique to *Ulva*,

while the pectin-lyase-kinase combination is found only in the multicellular algae *Ulva*, *Klebsormidium* and *Ectocarpus.* It is possible that *Ulva* Kringle-TM-kinase gene families arose via divergent evolution from a common ancestor, based on sequence similarity of family members and the close proximity of some family members on single DNA scaffolds. The pectin lyase-TM-kinase family proteins, on the other hand, are more divergent in sequence and structure (including kinase-TM-pectin lyase proteins) and are more likely to have arisen by dynamic gene fusions or conversions.

In addition to containing extracellular protein domains that are linked to intracellular kinases, *Ulva* also shows a significantly enriched diversity of the protein domains associated with the extracellular matrix (ECM) and cell surface, relative to its sequenced sister taxa. Notable examples of these enriched ECM-associated domains include Scavenger Receptor and Cysteine Rich domain (SRCR) proteins (IPR001190, IPR017448), which are absent from land plants but present in animals and Volvocales [34]. In Metazoans, the SRCR proteins have diverse roles that include the recognition of the pathogen-associated molecular patterns (PAMPs) that mediate bacterial interactions [35], and possibly encompass an early evolutionary role in cell-cell recognition or aggregation [36]. The germin (IPR001929) and RmlC-Like cupin domain folds (IPR011051) are also among the ECM-associated domains and occur ubiquitously in streptophytes, where they are linked to the regulation of cell wall properties such as extensibility and defense [37].

The gametolysin/MEROPS peptidase family M11 of VMPs (*Volvox* matrix metalloproteases) (IPR008752), similar to mammalian collagenases and crucial to ECM remodeling (Hallman et al. 2001), is enriched in *Volvox* relative to *Chlamydomonas* [6]. While this M11 class of peptidases is itself absent in *Ulva*, the related MEROPS M8/leishmanolysin peptidase domain (IPR001577) is 23-fold enriched compared to *Chlamydomonas.* Additionally, 28 collagen triple helix repeat (IPR008160) proteins, with predicted extracellular locations, have been detected, each with a G-N/D-E repeat, rather than the more usual G-P-Hyp, which suggests the presence of collagen-like innovations in the ECM of *Ulva*.

Phytohormones

The processes of growth and development in land plants are modulated by plant hormones, and experimental studies have suggested that such hormones may also be involved in *Ulva* morphogenesis to give a blade- or tube-like thallus [38]. However, because bacteria can produce plant hormones, it has not been clarified whether hormones previously detected in *Ulva* arose from the alga or from its associated bacteria. To add to the confusion, the *Roseovarius*-factor resembles a cytokinin and the *Maribacter*-factor

acts in a similar fashion to auxin, so bacteria-derived compounds may contribute to the development of the multicellular *Ulva* thallus [11]. We therefore investigated whether biosynthesis pathways for known phytohormones are present in *Ulva* and/or the associated bacteria and tested whether phytohormones could replace the bacterial morphogenetic factors. Homologs of plant hormone biosynthesis genes provided circumstantial evidence for the biosynthesis of abscisic acid (ABA), ethylene (ET) and salicylic acid (SA) in *Ulva* (Figure 6, Data S1). Corroborating these results, we found that both xenic and axenic *Ulva* produce not just ABA, ET and SA, but also auxin (IAA) and gibberellin (GA_3) (Table S4). Measurements of IAA in axenic cultures are more difficult to reconcile with the *Ulva* gene content, as little or no evidence was found for indole-3-pyruvic acid (IPA), tryptamine (TAM), and indole-3-acetamide (IAM) pathways. In *Ulva*, therefore, biosynthesis of IAA seems most likely to involve a pathway through indole-3-acetaldoxime and indole-3-acetaldehyde, catalysed by the AMI1 and AAO1 genes, respectively, that are found in the *Ulva* genome. The presence of GA_3 remains equivocal. In line with earlier findings [39], traces of GA_3 could be detected in both axenic and xenic slender cultures, but not in wildtype. At present, it is also unclear how to reconcile these traces of GA_3 with the gene content of *Ulva*, since the enzymes known to mediate the biosynthesis of the GA_3 precursor, *ent*-kaurene, in the plastid (CPP SYNTHASE, ENT-KAURENE SYNTHASE) are missing from the genome. We also found that the associated bacteria *Roseovarius* sp. MS2 and *Maribacter* sp. MS6 are able to produce IAA, GA_3, ET, SA and the cytokinins (iP) (Table S4). Despite this strong evidence for the synthesis of known phytohormones by *Ulva* and its associated bacteria, we found that none of the tested hormones were able to trigger growth or development of gametes and young propagules in standardized bioassays (Figure S3). And, while the *Ulva* genome contained candidate genes for phytohormone biosynthesis, the corresponding angiosperm receptors were absent, corroborating the comparative genomic analyses by Wang et al. [40]. These authors found little evidence for the emergence of homologous plant hormone signaling pathways outside the charophyte lineage. Our findings, however, do not preclude hormonal or other functions for these products, using different pathways and interdependencies, as demonstrated for diatoms [41].

Macroalgal-bacterial interactions

Ulva relies on interactions with bacteria for both the settlement its zoospores [42] and the morphogenesis of its thallus [10, 11]. This close association with bacteria, however, does not seem to have resulted in significantly higher levels of horizontal gene transfer (HGT). Using a phylogenomic pipeline to investigate the extent of HGT in the *Ulva* genome, we found 13 well-supported cases of HGT of prokaryotic origin

(Table S5). Although this number is not exceptionally high, it is remarkable that detected HGT events are followed by gene family expansion more than half the time. The most striking case is present by Haem peroxidases belonging to the peroxidase-cyclooxygenase superfamily, of which *Ulva mutabilis* has 36 copies that arose following a single HGT event. Peroxidases, which are involved in scavenging H_2O_2, are part of *Ulva*'s antioxidant machinery and help it to cope with the environmental challenges common to intertidal habitats, such as excessive light, hypersalinity and dehydration [16, 43]. Peroxidases may also have important functions in cell wall modification [44]; in plants, they have been demonstrated to anchor themselves to pectins and cross-link extensins, resulting in a stiffening of the primary cell wall. It is thus noteworthy that genes coding for extensins and pectin-like polysaccharides are very prominent in the *Ulva* genome (Data S1). Predictions of localization suggest that at least 21 of these peroxidases are extracellular, while a further 5 are predicted to be targeted to the mitochondrion.

A particularly interesting *Ulva*-bacterial interaction may involve the biosynthesis of siderophores for iron uptake. Pilot studies have recently revealed that bacteria tightly associated with *U. mutabilis* release unknown organic ligands that are capable of forming complexes with iron [45]. Secreted microbial siderophores become, therefore, part of the organic matter in the chemosphere and contribute to the recruitment of iron within the tripartite community of *U. mutabilis*. It was hypothesized that *Ulva* uses siderophores as public goods within a bacterial-algal mutualism, where heterotrophic bacteria are fed by the algae through the release of carbon sources. Genome mining for iron uptake genes suggests that *Ulva* can acquire iron, maintained in solution and made bioavailable by bacterial siderophores, via reduction by a ferric chelate reductase, re-oxidation by a multicopper ferroxidase and finally uptake by a transferrin-like protein (Data S1).

Dimethylsulfoniopropionate synthesis

Dimethylsulfoniopropionate (DMSP) has been identified as a major metabolite in *Ulva*. In addition to its role as an osmolyte and cryoprotectant, DMSP also appears to play a direct role in cross-kingdom signalling between *Ulva* and its associated bacteria [46]. *Roseovarius* sp. strain MS2 is chemotactically attracted to DMSP. Bacteria use the algal DMSP signal as a mechanism for detecting the presence of a photoautotrophic organism releasing various carbon sources such as glycerol [46]. In turn, *Roseovarius* sp. promotes the growth of *Ulva* by a cytokinine-like substance. DMSP works thus as an important

chemical mediator for macroalgal-bacteria interactions [46]. Comparative genome analysis reveals genes with a putative function in biosynthesis, catabolism and transport of DMSP.

The biosynthesis of dimethylsulfoniopropionate (DMSP) in *Ulva* has already been demonstrated to follow an entirely different route from that in higher plants [47]. In a four-step pathway, methionine is first transaminated to 4-methylthio-2-oxobutyrate (MTOB), then reduced to 4-methylthio-2-hydroxybutyrate (MTHB), and then methylated to give 4-dimethylsulfonio-2-hydroxy-butyrate (DMSHB) (Figure 7A). Fourthly, and finally, the oxidative decarboxylation of DMSHB yields DMSP. The methylation of MTHB to DMSHB is only known to occur in association with DMSP synthesis and is thus considered to be the pathway's committing step. The *DSYB* gene, a eukaryotic homolog of bacterial *dsyB*, has recently been shown to mediate the methylation of MTHB to DMSHB in several eukaryotic DMSP-producing algae [48], but no *DSYB* homolog could be found in either the *Ulva* genome, or in the genomes of the DMSP-producing diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* [49], which indicates that multiple biosynthesis pathways exist in eukaryotes. We can, however, locate an *Ulva* S-adenosyl-L-methionine-dependent methyltransferase (UM036_0102), which is homologous to a candidate methyltransferase in *Fragilariopsis cylindrus* (CCMP 1102, accession 212856a) that is known to be involved in DMSP production [50]. Homologs of UM036_0102 are notably absent from non-DMSP-producing Chlorophyceae and Trebouxiophyceae and expression analysis reveals that UM036_0102 is upregulated 7.0-fold (± 1.2 SD) at low temperatures (8°C) (Figure 7B). These expression values correlate with DMSP concentrations, which are significantly higher at 8°C (5.75 ± 1.8 mg / g FW) than at 18°C (3.14 ± 1.05 mg / g FW; Student's t-Test $p < 0.001$, $n = 12$), consistent with a potential role for DMSP as a cryoprotectant. The expression of UM052_056, a SAM methyltransferase with homologs in volvocine algae which was included as a control, was not increased. The combination of comparative analyses and expression data make UM036_0102 a credible candidate for the enzyme that mediates the methylation step in the biosynthesis of DMSP, as an alternative to *DSYB* in eukaryotes.

Also relevant when considering DMSP synthesis are four putative BCCT-type family symporters/antiporters (IPR000060), which have been demonstrated to be involved in the import of DMSP in several bacteria [51]. In *Ulva,* which shifts from DMSP synthesis towards DMSP uptake under conditions of sulfur deficiency, a putative DMSP transporter has been shown to be a Na^+/DMSP symporter, which points towards BCCT transporters [52]. Homologs of the putative *Ulva* BCCT transporters are also present in diatoms, some marine prasinophytes and opisthokonts, but are notably absent from the fresh water Trebouxiophyceae and Chlorophyceae. Two out of four BCCT transporters are significantly

upregulated at low temperatures (UM033_147 and 150), while one is significantly down-regulated (UM033_146) (Figure 7B).

Finally, the *Ulva* genome encodes two copies of a DMSP lyase, the enzyme responsible for forming DMS from DMSP (Figure 7A), which was originally identified in *Emiliania huxleyi* [53]. *Ulva* may thus regulate the chemoattraction of bacteria by the *de novo* synthesis of DMSP and its decomposition to DMS through its endogenous DMSP lyase. The production of DMSP and DMS by *Ulva* [47], and the ecological implications for the natural sulfur cycle and climate regulation (e.g. cloud condensation), are an important topic of research. The involvement of a DMSP lyase in the production of DMS has been demonstrated in the past [54], but the enzymes remained unidentified. Furthermore, the production of DMS by *Ulva* rather than by associated bacteria has only recently been unequivocally demonstrated [46]. Phylogenetic analyses resolved two DMSP lyases that are present in *U. mutabilis* in a clade with several haptophytes (including *Alma3*, *6* and *7* of *Emiliana huxleyi*) and dinoflagellates but also with the lyases of the scleractinian coral *Acropora* (Figure 7C). The most likely explanation for the presence of the Alma DMSP lyases in *Ulva*, the first reported in any green algal genome, would therefore be through HGT from the chromalveolate lineage rather than from bacteria. Of the two *Ulva* DMSP lyase copies, UM030_0039 was 13-fold downregulated at low temperatures (Figure 7B), while the expression of UM021_0036 did not vary significantly across treatments, which may indicate a function disconnected from temperature-responsive DMS formation.

Conclusion

Genome analysis of *Ulva mutabilis* reveals several key insights into the biology of green seaweeds. Most importantly, we demonstrate that *Ulva*'s long independent evolution from a unicellular and freshwater ancestral chlorophyte has resulted in marked changes in the genes underlying the most important aspects of the species' biology (e.g. cell cycle control). Although the unicellular–multicellular transition in volvocine algae relies on substantially different proximate causes, such as co-option of the retinoblastoma cell cycle pathway, interesting parallels can be drawn between expansions in the volvocine algae and the expansions of *Ulva*'s ECM-related gene families, especially given their putative role in environmental signaling. Additional ongoing advances in the precision and efficiency of reverse genetic and genome-editing techniques, combined with *Ulva*'s experimental tractability and ease of clonal haploid growth under laboratory conditions, make *Ulva* an exciting model organism for the study of the marine green

seaweeds. *Ulva* is increasingly used as a crop in seaweed aquaculture, and its sustainable and biosecure exploitation and domestication will benefit from the identification of ecotypic genetic variation; most immediately, in identifying the traits that underlie bloom formation. For example, the comparison of bloom- and non-bloom-forming *Ulva* species may assist our understanding of the molecular mechanisms that underpin growth and reproduction in response to environmental conditions. Furthermore, expanding comparative genomics to giant-celled green seaweeds (e.g. *Acetabularia*, *Caulerpa*, and *Cladophora*) has the potential to shed light on a range of curious macroscopic organismal architectures that do not rely on multicellularity for morphological patterning. Lastly, the reliance of *Ulva* on bacterial cues for growth and morphogenesis makes the species an exciting model to study the evolution of cross-kingdom signaling in the marine environment.

Acknowledgments

Research support was provided by NERC grant NBAF925 and BBSRC grant BB:K020552 (to C.A.M. and J.H.B.); UGent Special Research Fund BOF/01J04813 (to O.D.C., K.V.P.), BOF/01SC2316 (to X.L.) with infrastructure funded by EMBRC Belgium - FWO project GOH3817N (to O.D.C.); EU Horizon2020 Marie Curie ITN ALFF-Project 642575 (to C.M.G., O.D.C., T.W., G.C., Y.V.D.P.); EU COST Action FA1406 Phycomorph (to B.C., J.C., J.H.B., O.D.C., S.A.R. and T.W.); German Research Foundation CRC ChemBioSys 1127 (to T.W., J.Bo, M.K., G.C., L.L.); Postdoctoral Fellowship Grant of the Research Foundation – Flanders (to J.B. - Project 12T3418N); BBSRC-funded MIBTP PhD rotation project to R.C. and J.C.; United States Department of Energy DE-EE0003373/001 (to D.B.); National Science Foundation IGERT for Renewable and Sustainable fuels program 0903675 (to F.F.); M.H. is Research Associate of the FNRS. We thank Dr. Severin Sasso (University Jena) for providing T.W. with the Real-Time PCR Detection System.

Author Contributions

J.H.B., O.D.C., T.W., K.V., L.S. and Y.V.D.P. designed research; O.D.C, S.M.K., K.A.B., J.Bl. F.F., M.K., E.V., L.V., E.A., J.Bo., G.C., B.C., R.C., A.D.C., N.F.P., C.M.G., M.H., L.L., F.L., X.L., Z.A.P., M.V.B., P.I.K.W., D.B., J.C.C., S.A.R., D.V.D.S., L.S., T.W. and J.H.B. performed research; S.M.K., K.A.B., M.K., E.V., L.V., E.A., G.C., L.L., X.L., M.V.B., N.F.P., P.I.K.W., L.S., J.H.B. contributed new reagents/analytic tools; O.D.C, S.M.K., K.A.B., J.B., F.F., M.K., E.V., L.V., B.C., R.C., A.D.C., M.H., F.L., Z.A.P., J.A.R., D.B., J.C.C., S.A.R., D.V.D.S., A.V., L.S.,

K.V.P., T.W. and J.H.B. analyzed data; A.D.C., J.B., G.C., S.D., L.L., L.V., and X.L. prepared samples; and O.D.C., S.M.K., K.A.B, F.L., C.A.M., D.B., J.C.C., S.A.R., D.V.D.S., A.V., L.S., K.V., T.W., J.H.B., K.V., Y.V.D.P. wrote the paper.

Declaration Of Interests

The authors declare no competing interests.

References

1. Sebé-Pedrós, A., Degnan, B.M., and Ruiz-Trillo, I. (2017). The origin of Metazoa: a unicellular perspective. Nat. Rev. Genet. *18*, 498-512.
2. Rensing, S.A. (2016). (Why) Does Evolution Favour Embryogenesis? Trends Plant Sci. *21*, 562-573.
3. Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., and Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nature Plants *1*.
4. Umen, J.G., and Olson, B.J.S.C. (2012). Genomics of Volvocine Algae. In Genomic insights into the biology of algae, Volume 64, G. Piganeau, ed. (Academic Press), pp. 185-243.
5. Featherston, J., Arakaki, Y., Hanschen, E.R., Ferris, P.J., Michod, R.E., Olson, B.J.S.C., Nozaki, H., and Durand, P.M. (2017). The 4-celled *Tetrabaena socialis* nuclear genome reveals the essential components for genetic control of cell number at the origin of multicellularity in the volvocine lineage. Mol. Biol. Evol., 10.1093/molbev/msx1332.
6. Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science *329*, 223-226.
7. Hanschen, E.R., Marriage, T.N., Ferris, P.J., Hamaji, T., Toyoda, A., Fujiyama, A., Neme, R., Noguchi, H., Minakuchi, Y., Suzuki, M., et al. (2016). The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. Nat. Commun. *7*, 11370.
8. Leliaert, F., Verbruggen, H., Herron, M., Moreau, H., Smith, D.R., Delwiche, C., and De Clerck, O. (2012). Phylogeny and molecular evolution of the green algae. Crit. Rev. Plant Sci. *31*, 1-46.
9. Mine, I., Menzel, D., and Okuda, K. (2008). Morphogenesis in giant-celled algae. In International Review of Cell and Molecular Biology, Volume 266, K.W. Jeon, ed., pp. 37-83.
10. Wichard, T., Charrier, B., Mineur, F., Bothwell, J.H., De Clerck, O., and Coates, J.C. (2015). The green seaweed *Ulva*: a model system to study morphogenesis. Front. Plant Sci. *6*, 72.
11. Spoerner, M., Wichard, T., Bachhuber, T., Stratmann, J., and Oertel, W. (2012). Growth and thallus morphogenesis of *Ulva mutabilis* (Chlorophyta) depends on a combination of two bacterial species excreting regulatory factors. J. Phycol. *48*, 1433-1447.
12. Matsuo, Y., Imagawa, H., Nishizawa, M., and Shizuri, Y. (2005). Isolation of an algal morphogenesis inducer from a marine bacterium. Science *307*, 1598-1598.
13. Oertel, W., Wichard, T., and Weissgerber, A. (2015). Transformation of *Ulva mutabilis* (Chlorophyta) by vector plasmids integrating into the genome. J. Phycol. *51*, 963-979.
14. Smetacek, V., and Zingone, A. (2013). Green and golden seaweed tides on the rise. Nature *504*, 84-88.
15. Bolton, J.J., Cyrus, M.D., Brand, M.J., Joubert, M., and Macey, B.M. (2016). Why grow *Ulva*? Its potential role in the future of aquaculture. Perspectives in Phycology *3*, 113-120.
16. Zhang, X., Ye, N., Liang, C., Mou, S., Fan, X., Xu, J., Xu, D., and Zhuang, Z. (2012). De novo sequencing and analysis of the *Ulva linza* transcriptome to discover putative mechanisms associated with its successful colonization of coastal ecosystems. BMC Genomics *13*, 565.
17. Yamazaki, T., Ichihara, K., Suzuki, R., Oshima, K., Miyamura, S., Kuwano, K., Toyoda, A., Suzuki, Y., Sugano, S., Hattori, M., et al. (2017). Genomic structure and evolution of the mating type locus in the green seaweed *Ulva partita*. Scientific Reports *7*, 11679.
18. Ranjan, A., Townsley, B.T., Ichihashi, Y., Sinha, N.R., and Chitwood, D.H. (2015). An Intracellular Transcriptomic Atlas of the Giant Coenocyte *Caulerpa taxifolia*. PLoS Genet *11*, e1004900.

19. Føyn, B. (1958). Über die Sexualität und den Generationswechsel von *Ulva mutabilis*. Arch. Protistenkunde *102*, 473-480.
20. Løvlie, A. (1978). On the genetic control of cell cycles during morphogenesis in *Ulva mutabilis*. Dev. Biol. *64*, 164-177.
21. Harashima, H., Dissmeyer, N., and Schnittger, A. (2013). Cell cycle control across the eukaryotic kingdom. Trends Cell Biol. *23*, 345-356.
22. Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riaño-Pachón, D.M., Corrêa, L.G., Reski, R., Mueller-Roeber, B., and Rensing, S.A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol. Evol. *2*, 488-503.
23. Khanna, R., Kronmiller, B., Maszle, D.R., Coupland, G., Holm, M., Mizuno, T., and Wu, S.-H. (2009). The *Arabidopsis* B-box zinc finger family. The Plant Cell *21*, 3416-3420.
24. Cockram, J., Thiel, T., Steuernagel, B., Stein, N., Taudien, S., Bailey, P.C., and O'Sullivan, D.M. (2012). Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. PLoS One *7*, e45307.
25. Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G. (1995). The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell *80*, 847-857.
26. Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Más, P., Panda, S., Kreps, J.A., and Kay, S.A. (2000). Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. Science *289*, 768-771.
27. Liu, J., Shen, J., Xu, Y., Li, X., Xiao, J., and Xiong, L. (2016). Ghd2, a CONSTANS-like gene, confers drought sensitivity through regulation of senescence in rice. J. Exp. Bot. *67*, 5785-5798.
28. Corellou, F., Schwartz, C., Motta, J.-P., Sanchez, F., and Bouget, F.-Y. (2009). Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote *Ostreococcus*. The Plant Cell *21*, 3436-3449.
29. Serrano, G., Herrera-Palau, R., Romero, J.M., Serrano, A., Coupland, G., and Valverde, F. (2009). *Chlamydomonas* CONSTANS and the evolution of plant photoperiodic signaling. Curr. Biol. *19*, 359-368.
30. Lang, D., and Rensing, S.A. (2015). The Evolution of Transcriptional Regulation in the Viridiplantae and its Correlation with Morphological Complexity. In Evolutionary Transitions to Multicellular Life: Principles and mechanisms, I. Ruiz-Trillo and A.M. Nedelcu, eds. (Dordrecht: Springer Netherlands), pp. 301-333.
31. Holzinger, A., Herburger, K., Kaplan, F., and Lewis, L.A. (2015). Desiccation tolerance in the chlorophyte green alga *Ulva compressa*: does cell wall architecture contribute to ecological success? Planta *242*, 477-492.
32. Daher, F.B., and Braybrook, S.A. (2015). How to let go: pectin and plant cell adhesion. Front. Plant Sci. *6*, 523.
33. Green, J.L., Kuntz, S.G., and Sternberg, P.W. (2008). Ror receptor tyrosine kinases: orphans no more. Trends Cell Biol. *18*, 536-544.
34. Wheeler, G.L., Miranda-Saavedra, D., and Barton, G.J. (2008). Genome analysis of the unicellular green alga *Chlamydomonas reinhardtii* indicates an ancient evolutionary origin for key pattern recognition and cell-signaling protein families. Genetics *179*, 193-197.
35. Pereira, C.B., Bocková, M., Santos, R.F., Santos, A.M., de Araújo, M.M., Oliveira, L., Homola, J., and Carmo, A.M. (2016). The scavenger receptor SSc5D physically interacts with bacteria through the SRCR-containing N-terminal domain. Front. Immunol. *7*.
36. Bowdish, D.M., and Gordon, S. (2009). Conserved domains of the class A scavenger receptors: evolution and function. Immunol. Rev. *227*, 19-31.

37. Zimmermann, G., Bäumlein, H., Mock, H.-P., Himmelbach, A., and Schweizer, P. (2006). The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. Plant Physiol. *142*, 181-192.
38. Provasoli, L. (1958). Effect of plant hormones on *Ulva*. Biol. Bull. *114*, 375-384.
39. Gupta, V., Kumar, M., Brahmbhatt, H., Reddy, C., Seth, A., and Jha, B. (2011). Simultaneous determination of different endogenetic plant growth regulators in common green seaweeds using dispersive liquid–liquid microextraction method. Plant Physiol. Biochem. *49*, 1259-1263.
40. Wang, C., Liu, Y., Li, S.-S., and Han, G.-Z. (2015). Insights into the origin and evolution of the plant hormone signaling machinery. Plant Physiol. *167*, 872-886.
41. Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., Morales, R.L., Berthiaume, C.T., Parker, M.S., Djunaedi, B., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature *522*, 98-101.
42. Joint, I., Tait, K., Callow, M.E., Callow, J.A., Milton, D., Williams, P., and Cámara, M. (2002). Cell-to-cell communication across the prokaryote-eukaryote boundary. Science *298*, 1207-1207.
43. Sung, M.-S., Hsu, Y.-T., Hsu, Y.-T., Wu, T.-M., and Lee, T.-M. (2009). Hypersalinity and hydrogen peroxide upregulation of gene expression of antioxidant enzymes in Ulva fasciata against oxidative stress. Mar. Biotechnol. *11*, 199.
44. Passardi, F., Penel, C., and Dunand, C. (2004). Performing the paradoxical: how plant peroxidases modify the cell wall. Trends Plant Sci. *9*, 534-540.
45. Wichard, T. (2016). Identification of metallophores and organic ligands in the chemosphere of the marine macroalga *Ulva* (Chlorophyta) and at land-sea interfaces. Front. Mar. Sci. *3*.
46. Kessler, R.W., Weiss, A., Kuegler, S., Hermes, C., and Wichard, T. (2018). Macroalgal-bacterial interactions: Role of dimethylsulfoniopropionate in microbial gardening by *Ulva* (Chlorophyta). Mol. Ecol. *27*, 1808-1819.
47. Gage, D.A., Rhodes, D., Nolte, K.D., and Hicks, W.A. (1997). A new route for synthesis of dimethylsulphoniopropionate in marine algae. Nature *387*, 891.
48. Curson, A.R.J., Williams, B.T., Pinchbeck, B.J., Sims, L.P., Martínez, A.B., Rivera, P.P.L., Kumaresan, D., Mercadé, E., Spurgin, L.G., Carrión, O., et al. (2018). DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. Nat. Microbiol.
49. Curson, A.R., Liu, J., Martínez, A.B., Green, R.T., Chan, Y., Carrión, O., Williams, B.T., Zhang, S.-H., Yang, G.-P., and Page, P.C.B. (2017). Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. Nat. Microbiol. *2*, 17009.
50. Lyon, B.R., Lee, P.A., Bennett, J.M., DiTullio, G.R., and Janech, M.G. (2011). Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. Plant Physiol. *157*, 1926-1941.
51. Sun, L., Curson, A.R., Todd, J.D., and Johnston, A.W. (2012). Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. Biogeochemistry, 121-130.
52. Ito, T., Asano, Y., Tanaka, Y., and Takabe, T. (2011). Regulation of biosynthesis of dimethylsulfoniopropionate and its uptake in sterile mutant of *Ulva pertusa* (Chlorophyta). J. Phycol. *47*, 517-523.
53. Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D.S., and Vardi, A. (2015). Identification of the algal dimethyl sulfide–releasing enzyme: A missing link in the marine sulfur cycle. Science *348*, 1466-1469.
54. Steinke, M., and Kirst, G.O. (1996). Enzymatic cleavage of dimethylsulfoniopropionate (DMSP) in cell-free extracts of the marine macroalga *Enteromorpha clathrata* (Roth) Grev.(Ulvales, Chlorophyta). J. Exp. Mar. Biol. Ecol. *201*, 73-85.
55. Sterck, L., Billiau, K., Abeel, T., Rouze, P., and Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. Nat. Meth. *9*, 1041.

56. Wichard, T., and Oertel, W. (2010). Gametogenesis and gamete release of *Ulva mutabilis* and *Ulva lactuca* (Chlorophyta): Regulatory effects and chemical characterization of the "swarming inhibitor". J. Phycol. *46*, 248-259.
57. Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., and Sykes, S. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. U. S. A. *108*, 1513-1518.
58. Keller, O., Kollmar, M., Stanke, M., and Waack, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics *27*, 757-763.
59. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2015). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics *32*, 767-769.
60. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol., doi.org/10.1093/molbev/msx1319.
61. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722-736.
62. Veeckman, E., Ruttink, T., and Vandepoele, K. (2016). Are we there yet? Reliably estimating the completeness of plant genome sequences. The Plant Cell *28*, 1759-1768.
63. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Meth. *12*, 59.
64. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology *9*, R7.
65. Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.
66. Gouzy, J., Carrere, S., and Schiex, T. (2009). FrameDP: sensitive peptide detection on noisy matured sequences. Bioinformatics *25*, 670-671.
67. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Meth. *12*, 357.
68. Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., and Vandepoele, K. (2011). i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Res. *40*, e11-e11.
69. Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933-2935.
70. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., and Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236-1240.
71. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268-274.
72. Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics *29*, 2669-2677.
73. Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R., and Fondi, M. (2015). MeDuSa: a multi-draft based scaffolder. Bioinformatics *31*, 2443-2451.
74. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792-1797.

75. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology *16*, 157.
76. Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178-2189.
77. Felsenstein, J. (2005). PHYLIP: Phylogenetic inference program, version 3.6. University of Washington, Seattle.
78. Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H., Van de Peer, Y., Grimsley, N., and Piganeau, G. (2013). pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. Environ. Microbiol. *15*, 2147-2153.
79. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., and Young, S.K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one *9*, e112963.
80. Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics *30*, 1312-1313.
81. Smit, A., Hubley, R., and Green, P. (2013–2015). RepeatMasker Open-4.0 [Internet]. Institute for Systems Biology. Available from: www.repeatmasker.org.
82. Smit, A., and Hubley, R. (2008). RepeatModeler Open-1.0. Available fom www.repeatmasker.org.
83. Shao, M., and Kingsford, C. (2017). Accurate assembly of transcripts through phase-preserving graph decomposition. Nat. Biotechnol. *35*, 1167.
84. Simpson, J.T. (2014). Exploring genome characteristics and sequence quality without a reference. Bioinformatics *30*, 1228-1235.
85. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578-579.
86. Enright, A., Van Dongen, S., and Ouzounis, C. (2004). TribeMCL: An efficient algorithm for large scale detection of protein families. Retrieved November *21*, 2005.
87. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., and Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644.
88. Schattner, P., Brooks, A.N., and Lowe, T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. *33*, W686-W689.
89. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972-1973.
90. Hare, E.E., and Johnston, J.S. (2012). Genome size determination using flow cytometry of propidium iodide-stained nuclei. In Molecular methods for evolutionary genetics. (Springer), pp. 3-12.
91. Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., Wu, C.H., and Consortium, U. (2014). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics *31*, 926-932.
92. Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., and Barker, M. (2014). Data access for the 1,000 Plants (1KP) project. Gigascience *3*, 17.
93. Wilhelmsson, P.K.I., Mühlich, C., Ullrich, K.K., and Rensing, S.A. (2017). Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae. Genome Biol. Evol. *9*, 3384-3397.
94. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

95. Qiu, H., Price, D.C., Weber, A.P., Reeb, V., Yang, E.C., Lee, J.M., Kim, S.Y., Yoon, H.S., and Bhattacharya, D. (2013). Adaptation through horizontal gene transfer in the cryptoendolithic red alga Galdieria phlegrea. Curr. Biol. *23*, R865-R866.
96. Matsuura, T., Mori, I.C., Ikeda, Y., Hirayama, T., and Mikami, K. (2018). Comprehensive phytohormone quantification in the red alga Pyropia yezoensis by liquid chromatography–mass spectrometry. In Protocols for Macroalgae Research, B. Charrier, T. Wichard and C.R.K. Reddy, eds. (Boca Raton: CRC Press, Francis & Taylor Group).
97. Hu, Y., Depaepe, T., Smet, D., Hoyerova, K., Klíma, P., Cuypers, A., Cutler, S., Buyst, D., Morreel, K., and Boerjan, W. (2017). ACCERBATIN, a small molecule at the intersection of auxin and reactive oxygen species homeostasis with herbicidal properties. J. Exp. Bot. *68*, 4185-4203.
98. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*, research0034. 0031.
99. Team, R.C. (2018). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Austria, 2015. (ISBN 3-900051-07-0: URL http://www. R-project. org).

Figure and Table legends

Figure 1 | A. *Ulva mutabilis* external morphology of a blade (left) and tubular thallus (right). Both growth forms are part of the naturally occurring morphological variation (bar = 1 cm). B. Under axenic conditions, *Ulva* develops into a callus-like morphology without cell differentiation. Typical protrusions of the malformed exterior cell wall are visible (bar = 100 µm). C. *Ulva* can be easily cultured under standardized conditions associated with two bacterial strains, *Roseovarius* sp. MS2 and *Maribacter* sp. MS6, in batch cultures (bar = 1 cm). D. The surface of a blade with cells containing a single chloroplast and pyrenoid (differential interference contrast; bar = 15 µm). E. Transverse section of a typical *Ulva*-type blade with two cell layers (*U. rigida*; bar = 20 µm).

Figure 2 | Summary statistics for the *Ulva mutabilis* genome and comparison of genome size and number of protein-coding genes among green algal genomes. Color coding indicates classes.

Figure 3 | Predicted pattern of gain and loss of gene families during the evolution of green algae and land plants. The number of gene families acquired or lost (values indicated in blue along each branch in the tree) was estimated using the Dollo parsimony principle (see Methods). For each species, the total number of gene families, the number of orphans (genes that lack homologues in the eukaryotic data set) and the number of genes are indicated, as well as habitat and morphological characteristics. Maximum likelihood bootstrap values are indicated in black at each node.

Figure 4 | Comparative analysis of transcription-associated proteins. A. Heatmap of transcription factors comparing *Ulva* with a selection of green algae (*Bathycoccus prasinos*, bpr; *Chlamydomonas reinhardtii*, cre; *Chlorella variabilis*, CN64a; *Gonium pectorale*, gpe; *Micromonas pusilla*, mpu; *Micromonas* sp., m299; *Ostreococcus lucimarinus*, olu; *O. tauri*, ota and o809; *Coccomyxa subellipsoidea*, cvu; *Volvox carteri*, vca), streprophytes (*Klebsormidium nitens*, kni), land plants (*Arabidopsis thaliana*, ath; *Oryza sativa*, osa; *Physcomitrella patens*, ppa) and red algae (*Chondrus cripus*, ccr; *Cyanidioschyzon merolae*, cme) (Data S1). B. Maximum likelihood phylogeny (RAxML v8.2.8) of CO-like transcription factors which are expanded in *Ulva*. Roman numbers refer to the classification as in Khanna et al. [23]. C. Examples of tandem distributions of *Ulva* CO-like genes (containing a CCT and B-box domain), and genes containing either a CCT or B-box domain on contig 003, 053 and 154.

Figure 5 | Comparative analysis of enriched and depleted InterPro domains in *Ulva mutabilis* relative to *Chlamydomonas reinhardtii, Volvox carteri* and/or *Gonium pectorale* (Fisher's exact test, FDR corrected P-value < 0.05). Significant differences are denoted with squares if significant in *Ulva* and *Caulerpa*, circles if significant in *Ulva* only. Z-scores represent the number of IPR hits normalized by the total number of hits per species. Abbreviations see Table S7.

Figure 6 | Overview of phytohormone biosynthesis pathways and distribution of the biosynthetic enzymes across the Streptophyta and Chlorophyta lineages. (A) Current models of the biosynthesis pathways for the phytohormones ABA (abscisic acid), BR (brassinosteroids), CK (cytokinins), ET (ethylene), GA (gibberellins), IAA (auxin), JA (jasmonic acid), SA (salicylic acid), and SL (strigolactones). (B) Presence of putative homologs/orthologs of the main biosynthetic enzymes in 20 species based on the pico-PLAZA gene families and subfamilies. Phytohormones shown in green rectangles were identified in both axenic and non-axenic cultures of *Ulva mutabilis* (Table S5). Abbreviations: ZEP, zeaxanthin epoxidase; NCED, nine-cis-epoxycarotenoid deoxygenate; SDR, short-chain dehydrogenase reductase; AAO, aldehyde oxidase; DET2, deetiolated2 (steroid 5alfa reductase); DWF4, dwarf4

(CYP90B); CPD, constitutive photomorphogenic dwarf (CYP90A); IPT, isopentenyl transferase; LOG, lonely guy (lysine decarboxylase); SAMS, s-adenosyl methionine synthetase; ACS, ACC synthase; ACO, ACC oxidase; CPS, CDP/*ent*-kaurene synthase; KO, *ent*-kaurene oxidase; KAO, *ent*-kaurenoic acid oxidase; TAA, tryptophan aminotransferase; YUC, YUCCA (flavin monooxygenase); AMI, amidase; NIT, nitrilase; LOX, lipoxygenase; AOS, allene oxide synthase; AOC, allene oxide cyclase; OPR, oxo-phytodienoate reductase; ICS, isochorismate synthase; PAL, phenylalanine ammonia-lyase; D27, dwarf27 (all-trans/9-cis-B-carotene isomerase); CCD, carotenoid cleavage dioxygenase; MAX1, more axillary branches1 (CYP711A). For a complete list of abbreviations see Data S1.

Figure 7 | Biosynthesis, transport and catabolism of DMSP. A. Maintenance of DMSP concentration in *Ulva* is a combination of *de novo* synthesis by conversion of methionine [47], import from the environment putatively using BCCT transporters [52], and degradation into acrylate and DMS by DMSP lyases [53]. B. Expression analysis (qPCR) of the Alma DMSP lyase homologues, putative S-adenosyl-L-methionine-dependent methyltransferases and BCCT transporters in xenic and axenic *Ulva* at 18°C and 8°C. Results shown are based on 4 replicates, except UM030_0039 for which 8 replicates were used. Significant expression values (Fisher's post hoc test) are indicated for each gene. C. Maximum likelihood phylogeny (RAxML v8.2.8) of the DMSP lyase genes, indicating lateral gene transfer of the Alma gene from the 'chromalveolate' lineage.

Data S1. Annotated genes of *Ulva mutabilis* transcription associated genes, cell cycle related genes, protein kinases, phytohormone synthesis and signalling, cell wall synthesis, and iron uptake.

STAR Methods

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
TruSeq DNA PCR-Free Library Preparation Kit	Illumina	Catalog #: FC-121-3001
Nextera Mate Pair Sample Preparation Kit	Illumina	Catalog #: FC-132-1001
ScriptSeq v2 RNA-Seq Library Preparation Kit	Illumina	Catalog #: SSV21106
Deposited Data		
Ulva mutabilis: raw and analyzed sequence data	EMBL	BioProject: PRJEB25750
Online annotation platform	ORCAE [55]	ORCAE - http://bioinformatics.psb.ugent.be/orcae/overview/Ulvmu
Experimental Models: Organisms/Strains		
Ulva mutabilis Føyn (wildtype and slender strains)	[56]	Friedrich Schiller University Jena, Germany
Roseovarius sp. MS2	[11]	Friedrich Schiller University Jena, Germany
Maribacter sp. MS6	[11]	Friedrich Schiller University Jena, Germany
Software and Algorithms		
Genome size estimation	ALLPATHS-LG [57]	http://software.broadinstitute.org/allpaths-lg/blog/?page_id=12
Gene prediction	Augustus-PPX [58]	http://bioinf.uni-greifswald.de/augustus/
Assembly polishing	Arrow (v2.2.1) (SMRT Link v5.0.1.9585)	https://downloads.pacbcloud.com/public/software/installers/smrtlink_5.0.1.9585.zip
Gene prediction	BRAKER1 (v1.9) [59]	http://exon.gatech.edu/braker1.html
Genome completeness	BUSCO (v3.0.2b) [60]	https://gitlab.com/ezlab/busco
Genome assembly	Canu (v1.6) [61]	https://github.com/marbl/canu
Genome completeness	Core gene family analysis [62]	ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_02_5/coreGF
Protein alignment prior to gene prediction	Diamond (v0.9.18) [63]	https://github.com/bbuchfink/diamond
Gene prediction	EVidenceModeler [64]	https://evidencemodeler.github.io/

Gene prediction	Exonerate (v2.2.0) [65]	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
Reading frame correction	FrameDP (v1.2.2) [66]	https://iant.toulouse.inra.fr/FrameDP/
RNA-seq mapping	HISAT2 (v2.0.5) [67]	https://ccb.jhu.edu/software/hisat2/index.shtml
Detection of collinear regions	i-ADHoRe (v3.0.01) [68]	http://bioinformatics.psb.ugent.be/beg/tools/i-adhore30
Prediction of non-coding RNA	Infernal [69]	http://eddylab.org/infernal/
Functional annotation	InterProScan (v5.27-66) [70]	https://www.ebi.ac.uk/interpro/
Maximum likelihood tree calculation	IQtree (v1.4.3) [71]	http://www.iqtree.org/
Genome assembly	MaSuRCA (v3.2.3) [72]	https://github.com/alekseyzimin/masurca
Contig scaffolding	MEDUSA [73]	https://github.com/combogenomics/medusa
Sequence alignment	Muscle (v3.8.31) [74]	https://www.ebi.ac.uk/Tools/msa/muscle/
Gene family prediction	OrthoFinder (v2.1.2) [75]	https://github.com/davidemms/OrthoFinder
Gene family prediction	OrthoMCL [76]	http://orthomcl.org/orthomcl/
Gene prediction	PASA [64]	https://github.com/PASApipeline/PASApipeline
Long-read mapping	pbalign (v0.3.1) (SMRT Link v5.0.1.9585)	https://downloads.pacbcloud.com/public/software/installers/smrtlink_5.0.1.9585.zip
Gene family loss/gain analysis	PHYLIP (Dollop) [77]	http://evolution.genetics.washington.edu/phylip.html
Comparative genomics	Pico-PLAZA (v2.0) [78]	https://bioinformatics.psb.ugent.be/plaza/versions/pico-plaza/
Assembly polishing	Pilon (v1.20) [79]	https://github.com/broadinstitute/pilon
Maximum likelihood tree calculation	RAxML (v8.2.4) [80]	https://sco.h-its.org/exelixis/web/software/raxml/index.html
Repetitive DNA-motif masking	RepeatMasker (v4.0.7) [81]	https://github.com/rmhubley/RepeatMasker
Repetitive DNA-motif identification	RepeatModeler (1.0.8) [82]	https://github.com/rmhubley/RepeatModeler

Reference-based transcriptome assembly	Scallop (v0.10.2) [83]	https://github.com/Kingsford-Group/scallop
Genome size estimation	SGA PreQC [84]	https://github.com/jts/sga
Contig scaffolding	SSPACE (v3.0) [85]	https://github.com/nsoranzo/sspace_basic
Gene family prediction	TribeMCL [86]	https://micans.org/mcl/
Transcriptome assembly	Trinity v2.4.0 [87]	https://github.com/trinityrnaseq/trinityrnaseq
tRNA identification	tRNAscan-SE (v1.31) [88]	http://lowelab.ucsc.edu/tRNAscan-SE/
Alignment curation	TrimAI (v1.2) [89]	http://trimal.cgenomics.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Olivier De Clerck (Olivier.declerck@ugent.be).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The *U. mutabilis* strain sequenced, a wildtype gametophyte of mating type minus (wt-G(mt$^-$) (gametophyte); ([mt$^-$]; G/PS- swi$^+$; mut-; RS140$^+$; RS180$^+$)) was initially isolated from Ria Formosa, southern Portugal, by B. Føyn [13, 19]. An additional haploid strain, *slender*, a spontaneous mutant derived from the original collection, was selected to complement the available transcriptomes because of its fast-growing nature and the ease with which thalli can be induced for gamete formation. Strains of *U. mutabilis* are primarily maintained at the Friedrich-Schiller-Universität Jena. Gametophytes were raised parthenogenetically from unmated gametes in *Ulva* Culture Medium (UCM without antibiotics) in the presence of bacterial symbionts *Roseovarius* sp. MS2 (Genbank: EU359909) and *Maribacter* sp. MS6 (Genbank: EU359911) to secure normal thallus morphogenesis. Bacterial strains are cultivated in marine broth medium (Roth, Germany). Algae were cultured at 20°C, under long day light conditions (L/D 17:7) consisting of 60–120 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (50% GroLux, 50% day-light fluorescent tubes; OSRAM, München, Germany), without aeration.

METHOD DETAILS

DNA and RNA extraction and library construction

DNA for sequencing was prepared from axenic gametes. For mate pair libraries adult, non-axenic *Ulva* tissue was used. DNA extraction was performed using a CTAB protocol. The PacBio library was prepared using P6-C4.0 chemistry with size selection using a 0.75% cassette on a Blue Pippin instrument with a lower cutoff of 8 kb based on the Pacific Biosciences 20kb template protocol (ref 100-286-000-08). Libraries for short reads and mate pairs were constructed using the TruSeq DNA PCR-Free Library Preparation Kit and the Nextera Mate Pair Sample Preparation Kit, respectively (Illumina, San Diego, CA). Total RNA was isolated from vegetative adult tissue, tissue in the process of gamete formation as well as gametes using the RNeasy Mini Kit (Qiagen). cDNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit.

Genome sequencing and assembly

Sequencing. Genomic DNA was sheared to produce fragments of 350 to 550 bp and sequenced on a MiSeq2000 (2x250 bp PE reads). Mate-pair libraries were sequenced on an Illumina HiSeq 2500 (2x125 bp reads). The PacBio libraries were sequenced on a PacBio RSII instrument (five SMRTCells P6-C4 chemistry). RNA libraries were sequenced on one lane of Illumina HiSeq 2500 (2x125 bp PE reads) and one run of NextSeq 550 (1x150 bp SE reads). Table S6 presents an overview of the sequenced libraries.

Estimation of genome size. Both k-mer analysis and flow cytometry experiments were used to gauge the genome size of *Ulva mutabilis* (wild-type). Based on the concatenated Illumina paired-end libraries the genome size was estimated 93.6 Mbp by SGA PreQC [84] with default k-mer size 31 and 104.5 Mbp by the estimating process in ALLPATHS-LG [57] with k-mer size 25. For flow cytometry estimates, *Ulva mutabilis* nuclei were stained together with nuclei of the standard (*Arabidopsis thaliana*) and relative fluorescence was used to calculate the genome size as described by Hare & Johnston [90]. Fluorescence emission was collected using the S3e™ Cell Sorter (BIO-RAD). The flow cytometry measurements showed an estimated genome size of 100.2 ± 3.6 Mbp (mean ± standard error of four measurements).

Genome assembly. PacBio reads (6.9 Gbp) were assembled using Canu [61] resulting in a 98.4 Mbp assembly in 1,119 contigs. The 30x of the longest corrected reads from the Canu pipeline (N50 of 9.0 kbp) and the Illumina paired-end reads were used by MaSuRCA [72] to generate a hybrid assembly, resulting in an assembly of 108.1 Mbp, with scaffold N50 of 264.9 kbp and longest scaffold in 2.7 Mbp. A graph-based scaffolder MeDuSa [73] was used to scaffold the Canu contigs based on the MaSuRCA assembly, followed by SSPACE [85] scaffolding using all the mate-pair libraries. The super-scaffolds were first polished by PacBio reads using Arrow v2.2.1 (from SMRT Link v5.0.1) after mapping all the long reads by

pbalign v0.3.1 (from SMRT Link v5.0.1). The PacBio-polished scaffolds were further improved using the paired-end Illumina reads with Pilon [79]. To eliminate putative bacterial contamination super-scaffolds were searched against the NCBI nucleotide (nt) database using MegaBLAST with an (E-value < 1e-65).

***De novo* repeat finding and repeat masking**

A *de novo* repeat identification was performed with RepeatModeler (1.0.8) [82]. Unknown elements were screened with BlastX (E-value < 1e-5) against UniRef90 database [91] (subset Viridiplantae) and removed from the repeat library if necessary. The filtered *Ulva* repeat library was used by RepeatMasker (4.0.7) [81] to mask the repetitive elements in the assembly, which resulted in 34.7 Mbp (35.28%) of the genome masked.

Gene prediction

We applied EVidenceModeler [64] to predict gene models. The consensus gene models were reconciled using the models from *ab initio* and orthology-aided predictions, transcripts reconstructed from RNA-Seq, and homologous models derived from the protein alignments of the available public resource (Figure S5A, 5B). We used BRAKER1 v1.9 [59] to predict the gene models incorporating the RNA-Seq mapping results generated using HISAT2 v2.0.5 [67]. We further used Augustus v3.2.3 with the trained data from BRAKER1 and the protein profile extension to re-predict the gene models [58]. Protein profiles were generated by processing the missing family identified after gene family assignment using OrthoFinder v2.1.2 [75] with the following reference sequence from Phytozome v12.1: *Chlamydomonas reinhardtii* v5.5, *Coccomyxa subellipsoidea* C-169 v2.0, *Dunaliella salina* v1.0, *Gonium pectorale* (assembly ASM158458v1) and *Volvox carteri* v2.1. In addition to the *ab initio* prediction, the RNA-Seq data were also used to reconstruct the transcripts, which consisted of consensus transcripts predicted by Scallop v0.10.2 [83] and predicted coding regions of Trinity v2.4.0 [87] assemblies (both *de novo* and genome-guided) using PASA [64]. Spliced alignments of proteins from UniRef90 (with taxonomy ID 33090) and the FrameDP-corrected [66] and predicted proteins sequence of green algal transcriptomes in the oneKP project [92] were generated using Exonerate [65] seeded by Diamond [63]. We combined the aforementioned gene models with the alignments of proteins, annotation of repetitive elements to produce a consensus gene set using EVidenceModeler. Non-coding RNA and tRNA were identified using Infernal v1.1 [69] and tRNAscan-SE v1.31 [88], respectively. Gene models were functionally annotated using InterProScan 5.27-66 [70] and uploaded to the ORCAE platform [55], enabling members of the consortium to curate and manually

annotate. Transcription factors (TF) and transcriptional regulators (TR) were annotated by first screening the proteins for domains and then applying a domain-based rule set [22, 93].

Genome completeness

Completeness of the predicted *Ulva* gene space was evaluated using BUSCO [60] and the coreGF analysis [62]. Core gene families were defined as gene families shared among all *Chlorophyta* present in pico-PLAZA v. 2.0 [78]. In total 1,815 gene families were compared by sequence similarity to the *Ulva* gene models and the reference protein sequences. We assigned each *Ulva* gene model to a particular gene family based on the top 5 hits (E-value < 1e-5). Finally, a GF score was calculated as the sum of each core family identified, counted with a weight equal to one divided by the average family size. Hence, a GF score of 1 indicates that all the core gene families were identified, while a GF score of 0 indicates that no core gene families were found. The likelihood of the presence of each core gene family was calculated for *Ulva* gene models.

Comparative genomic analyses

For the comparative genomic analyses a custom version of Pico-Plaza [78] was built containing genomes and annotations of 32 eukaryotic species (Table S7). Following an 'all-versus-all' blastP [94] protein sequence similarity search, both TribeMCL v10-201 [86] and OrthoMCL v2.0 [76] were used to delineate gene families and subfamilies. Collinear regions (regions with conserved gene content and order) were detected using i-ADHoRe 3.0 [68] with the following settings: gf_type = TribeMCL gene families, alignment method = gg2, number of anchor point s=5, gap size = 30, cluster gap = 35, tandem gap = 30, q-value = 0.85, probability cut-off = 0.01, multiple hypothesis correction = FDR and level_2_only = false. The phylogenetic profile of TribeMCL gene families (excluding orphans) retrieved from pico-PLAZA and the inferred species tree topology were provided to reconstruct the most parsimonious gain and loss scenario for every gene family using the Dollop program from PHYLIP v3.69 [77]. Gene family losses and gains were further analysed by examining the associated Gene Ontology (GO) terms and InterPro domains. Functional information was retrieved using InterProScan [70].

From TribeMCL gene families, highly conserved families were defined as single copy gene families present in all 20 species (*U. mutabilis*, 3 Chlorophyceae, 6 Trebouxiophyceae, 6 prasinophytes and 4 Streptophyta). A concatenated alignment of 58 single copy genes (42,401 amino acid positions) was used to construct a phylogenetic tree using RAxML v8.2.8 [80] (model PROTGAMMAWAG, 100 bootstraps).

Horizontal gene transfer

To search for HGTs, a blast search *U. mutabilis* proteome (blastp, E-value < 10^{-5}) was carried out against a reduced RefSeq database complemented with data from several algal genomes [95]. The top 1,000 Blastp hits (sorted by bit score) from each query were parsed via custom scripts to extract ≤12 representatives from each phylum to create a taxonomically diverse sample. The blastp hits were re-ordered according to query-hit identity followed by the sampling of another set of representative sequences. The query sequence was then combined with the sets of sampled representative sequences following Qiu et al. [95]. A custom script was used to sort trees consisting of *U. mutabilis* that was nested among prokaryotes with at least 80% bootstrap supports. Candidate HGT sequences were then reanalyzed using IQtree v1.4.3 [71] with the built-in model selection function, and branch support estimated using ultrafast bootstrap (UFboot) with 1,500 bootstrap replicates (-bb 1,500). Last, we visually confirmed that the putative HGT genes were an integral part of long sequencing reads and flanked up and downstream by non-HGT genes.

Phytohormone bioassay and measurements

Endogenous phytohormones were extracted from thalli (3-weeks-old) and axenic cultures (12.5 mg·mL^{-1}) by the acidified polar solvent 80 % acetonitrile 1 % acetic acid in water (v:v) upon maceration through a tissue lyser (Qiagen, Germany). Compounds were purified to eliminate interferents by multiple steps of solid-phase extractions [96], and quantified by UHPLC-ESI-MS/MS (LC system equipped with a C18-Kinetex column (dimensions: 50 × 1.7 mm, Phenomenex, USA) and coupled to a Q Exactive Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, UK). Deuterium-labelled standards (d_5-IAA, d_6-ABA, d_4-SA; ChemIm Czech Republic) were used for quantification. The limit of detection was in the range of 0.8 - 8.0 fmol on the column, whereas the limit of quantification was in the range of 2 - 20 fmol on column depending on the phytohormone. Ethylene emanation was measured by laser photo-acoustics [97].

To test if morphogenesis of *Ulva* into a blade- or tube-like thallus may require specific phytohormones, gametes were inoculated with ABA, GA_3, IAA, JA, SA and ZEA at a final concentration of 10^{-6} and 10^{-9} mol/l. Control treatments consisted of inoculating *Ulva* gametes with bacteria (final optical density 1x10^{-4}). Development was monitored for 2 weeks.

Measurement of DMSP and qPCR.

Expression of DMSP lyases (UM030_0039, UM021_0036), putative S-adenosyl-L-methionine-dependent methyltransferases (UM052_056, UM036_0102) and BCCT transporters (UM033_146, UM033_147, UM033_150, UM033_150) was examined in xenic and axenic 4-5 week-old thalli grown at 8°C or 18°C using qPCR (Table S8). Expression values are based on at least 4 biological replicates. DMSP was measured in xenic thalli grown under both temperatures as outlined in Kessler et al. [46] (n = 12). RNA was extracted from *Ulva mutabilis* (slender strain) using the Spectrum Plant Total RNA Kit (Sigma), including the optional DNaseI digestion, and 500 ng was reverse-transcribed with the PrimeScript RT-PCR Kit (Clontech). qPCR reactions were performed using SYBR Green Master Mix (Clontech) on a BioRad CFX96 Real-Time PCR Detection System. Primers are listed. Transcript abundance was determined according to the Pfaffl method [98]. For normalization of expression levels, we used the reference genes UM008_0183 (Ubiquitin) and UM010_0003 (PP2A 65 kDa regulatory subunit A).

QUANTIFICATION AND STATISTICAL ANALYSIS

Enrichment and depletion of InterPro domains in *Ulva* and *Caulerpa* were tested using a Fisher's exact tests with a false discovery rate correction (Benjamini-Hochberg FDR method) of 0.05. Interpro domains, proven to be significantly depleted or expanded were grouped at Superfamily-level to reduce redundancy. Fisher's post hoc tests were used to determine significance of expression values of DMSP lyases (n = 4 or 8), methyltransferases (n = 4) and BCCT transporters (n = 4). All statistical tests were carried out in R [99]. Phylogenetic trees were reconstructed using either RAxML v8.2.8 [80], IQTree v1.4.3 [71].

DATA AND SOFTWARE AVAILABILITY

The assembled genome sequence, annotation and raw reads are accessible from ENA under BioProject number PRJEB25750. Further information on the *Ulva mutabilis* project is available via the Online Resource for Community Annotation of Eukaryotes (ORCAE) at http://bioinformatics.psb.ugent.be/orcae/.

Figure 1

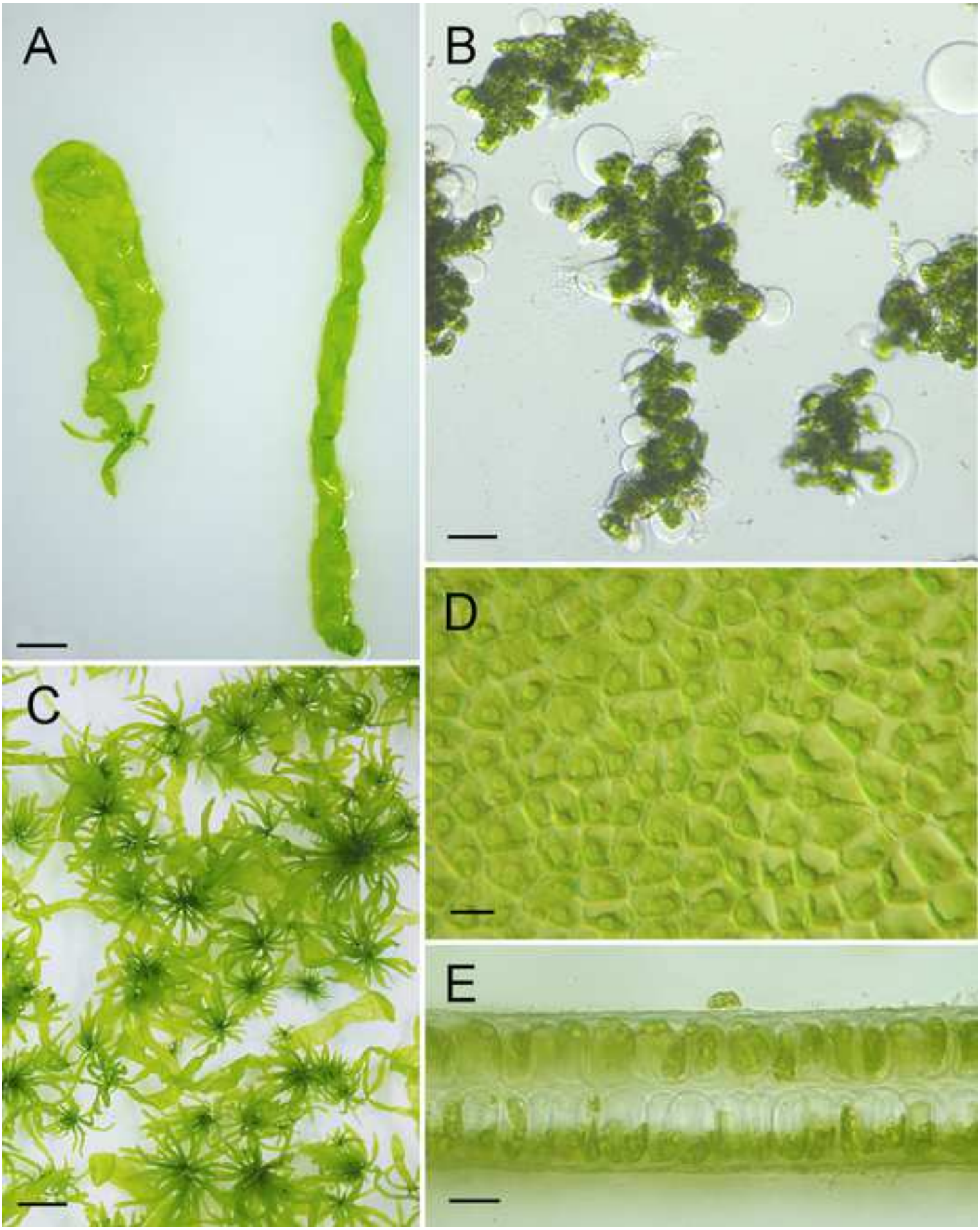


Figure 2

Ulva mutabilis: summary statistics

Genome size (Mbp)	98.5
Scaffold N50 (Mbp)	0.6
Percentage GC	57.2
Number of protein coding genes	12,924
Gene density (genes/Mb)	131.2
Average intron per gene	3.9
Average exon length (bp)	303.1
Average intron length (bp)	368.6
Percentage of genes with introns	85.0

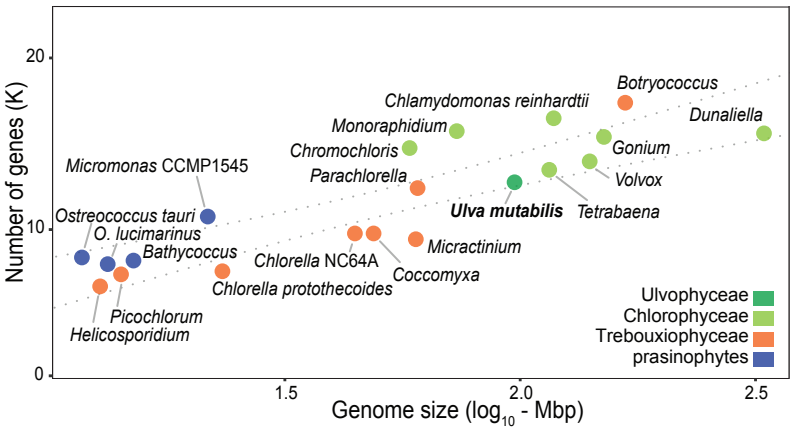


Figure 3

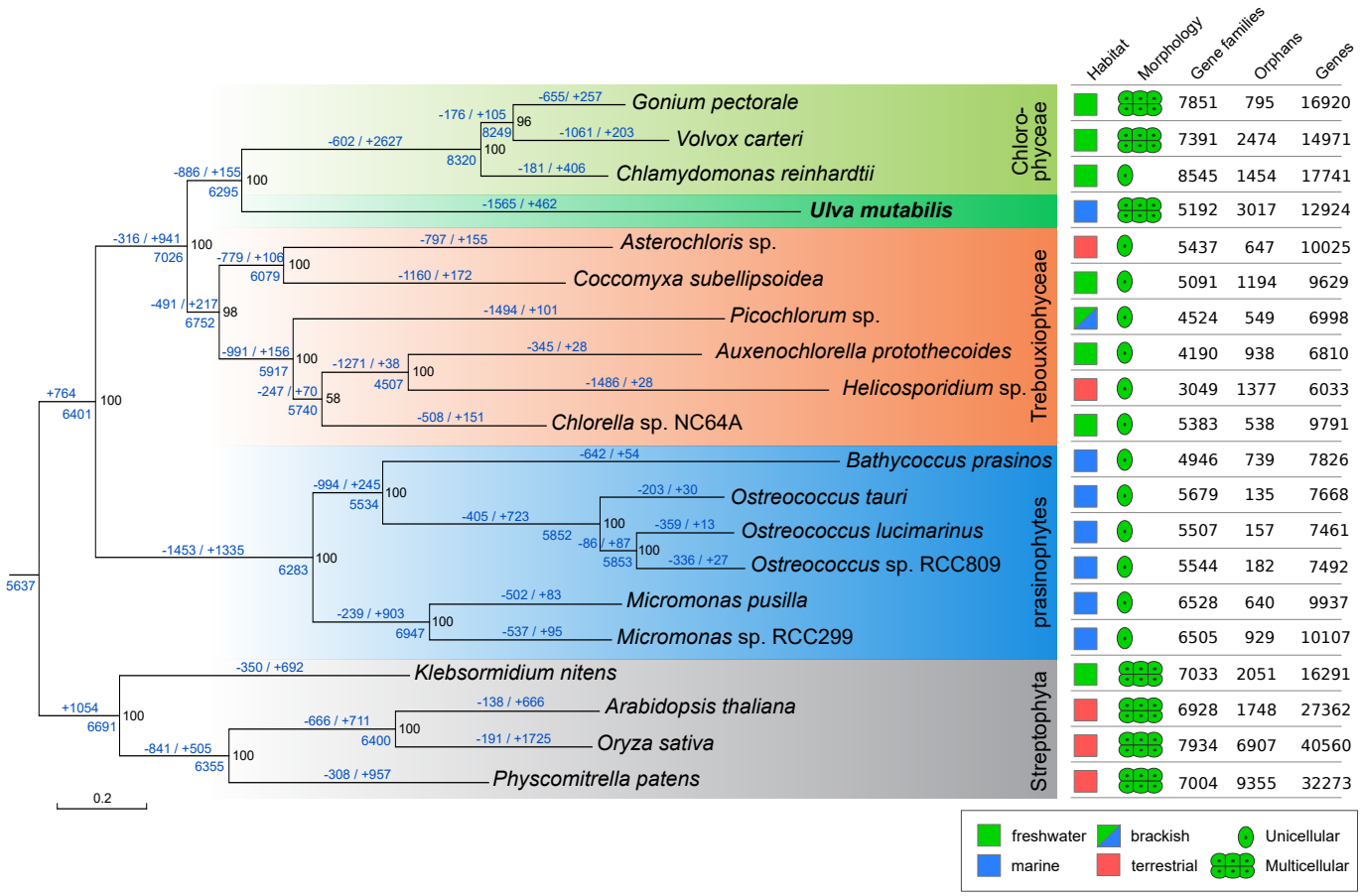

Habitat
Morphology
Gene families
Orphans
Genes
Chloro-phyceae
Trebouxiophyceae
prasinophytes
Streptophyta
Gonium pectorale
Volvox carteri
Chlamydomonas reinhardtii
Ulva mutabilis
Asterochloris sp.
Coccomyxa subellipsoidea
Picochlorum sp.
Auxenochlorella protothecoides
Helicosporidium sp.
Chlorella sp. NC64A
Bathycoccus prasinos
Ostreococcus tauri
Ostreococcus lucimarinus
Ostreococcus sp. RCC809
Micromonas pusilla
Micromonas sp. RCC299
Klebsormidium nitens
Arabidopsis thaliana
Oryza sativa
Physcomitrella patens
7851 795 16920
7391 2474 14971
8545 1454 17741
5192 3017 12924
5437 647 10025
5091 1194 9629
4524 549 6998
4190 938 6810
3049 1377 6033
5383 538 9791
4946 739 7826
5679 135 7668
5507 157 7461
5544 182 7492
6528 640 9937
6505 929 10107
7033 2051 16291
6928 1748 27362
7934 6907 40560
7004 9355 32273
0.2
freshwater
brackish
Unicellular
marine
terrestrial
Multicellular

Figure 4

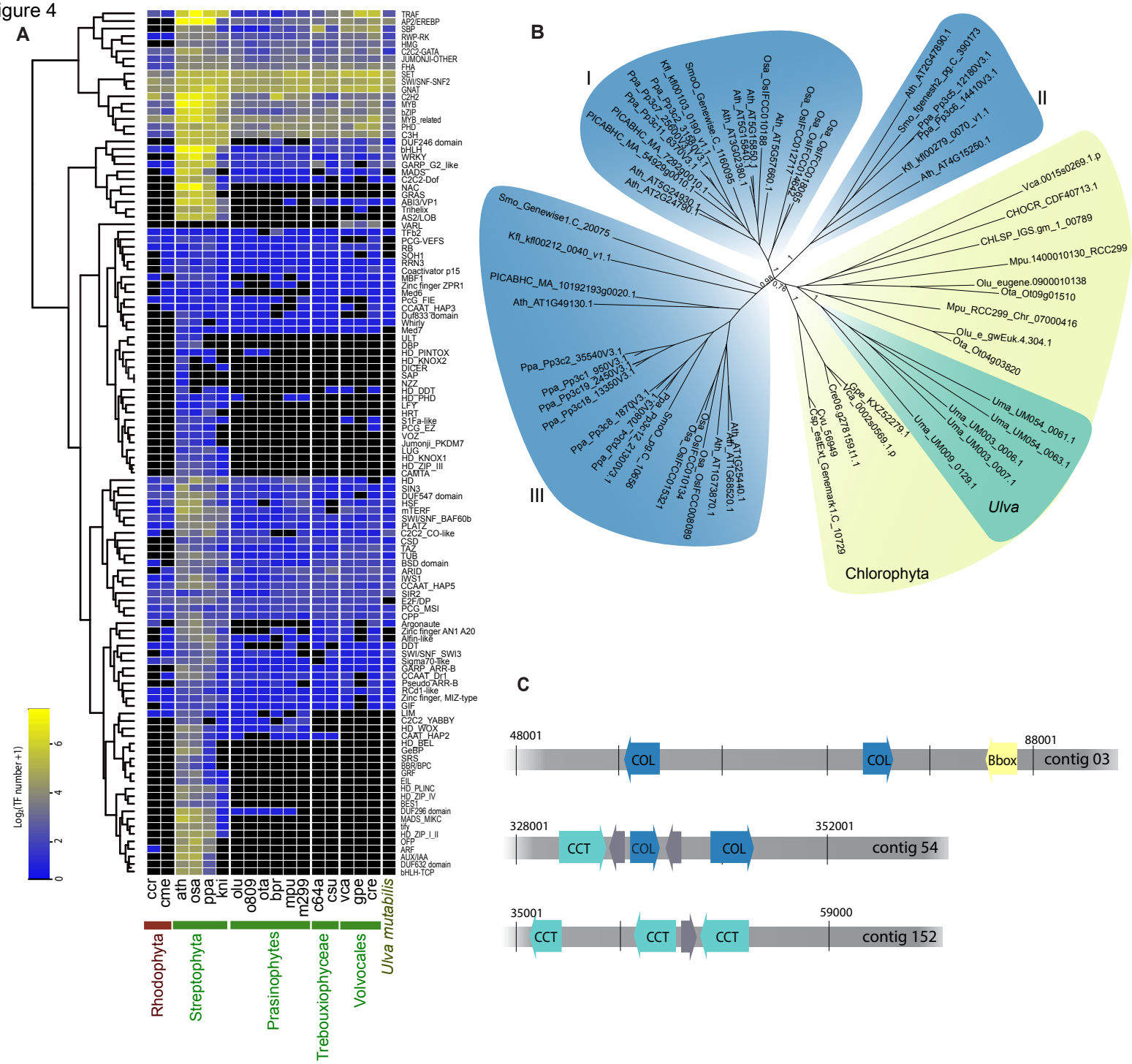

A
B
C
Log2(TF number + 1)
ccr cme ath osa ppa kni olu o809 ota bpr mpu m299 c64a csu vca gpe cre
Rhodophyta
Streptophyta
Prasinophytes
Trebouxiophyceae
Volvocales
Ulva mutabilis
I
II
III
Chlorophyta
Ulva
48001
88001
COL
COL
Bbox
contig 03
328001
352001
CCT
COL
COL
contig 54
35001
59000
CCT
CCT
CCT
contig 152

Figure 5

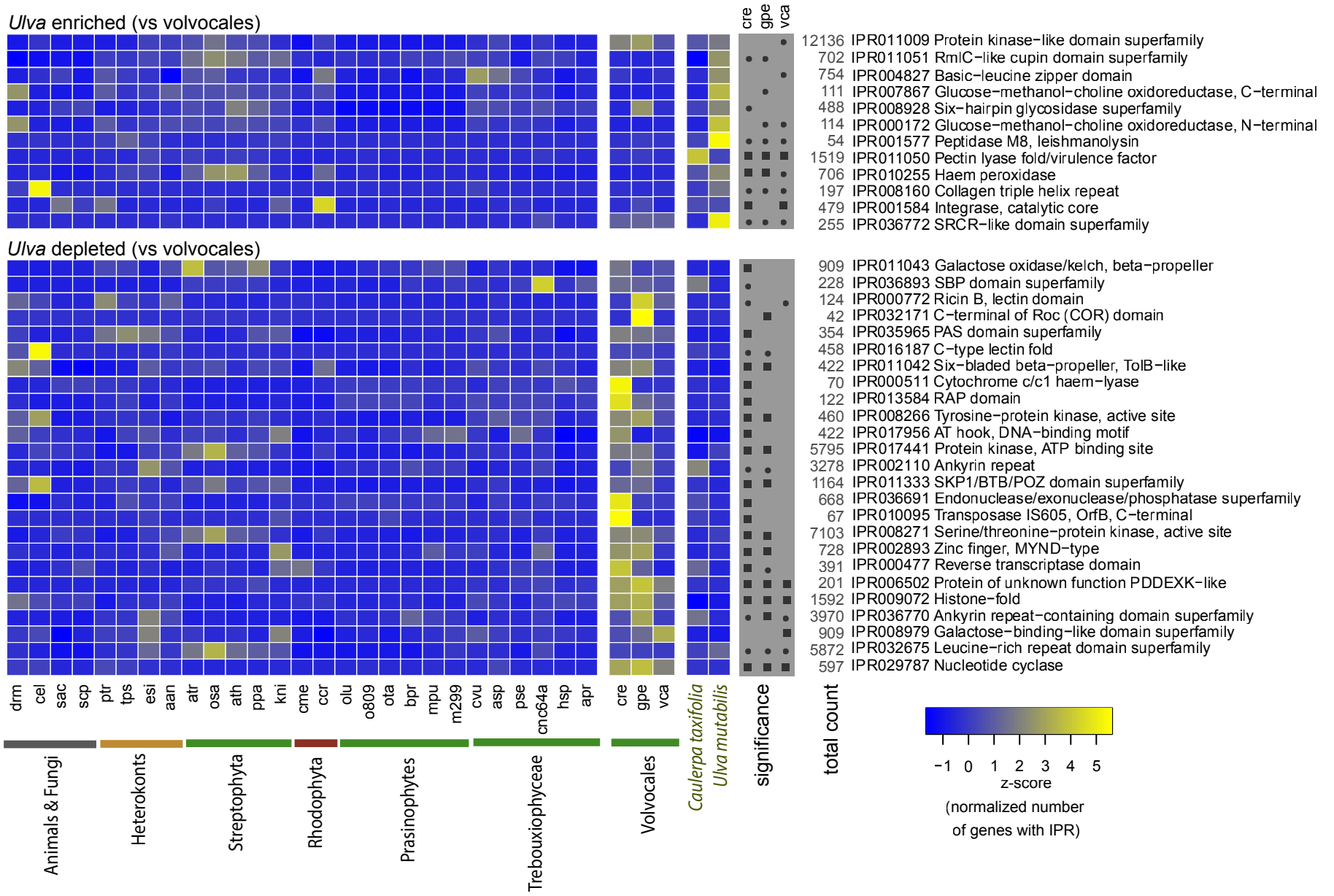

Ulva enriched (vs volvocales)
Ulva depleted (vs volvocales)
cre gpe vca
12136 IPR011009 Protein kinase−like domain superfamily
702 IPR011051 RmlC−like cupin domain superfamily
754 IPR004827 Basic−leucine zipper domain
111 IPR007867 Glucose−methanol−choline oxidoreductase, C−terminal
488 IPR008928 Six−hairpin glycosidase superfamily
114 IPR000172 Glucose−methanol−choline oxidoreductase, N−terminal
54 IPR001577 Peptidase M8, leishmanolysin
1519 IPR011050 Pectin lyase fold/virulence factor
706 IPR010255 Haem peroxidase
197 IPR008160 Collagen triple helix repeat
479 IPR001584 Integrase, catalytic core
255 IPR036772 SRCR−like domain superfamily
909 IPR011043 Galactose oxidase/kelch, beta−propeller
228 IPR036893 SBP domain superfamily
124 IPR000772 Ricin B, lectin domain
42 IPR032171 C−terminal of Roc (COR) domain
354 IPR035965 PAS domain superfamily
458 IPR016187 C−type lectin fold
422 IPR011042 Six−bladed beta−propeller, TolB−like
70 IPR000511 Cytochrome c/c1 haem−lyase
122 IPR013584 RAP domain
460 IPR008266 Tyrosine−protein kinase, active site
422 IPR017956 AT hook, DNA−binding motif
5795 IPR017441 Protein kinase, ATP binding site
3278 IPR002110 Ankyrin repeat
1164 IPR011333 SKP1/BTB/POZ domain superfamily
668 IPR036691 Endonuclease/exonuclease/phosphatase superfamily
67 IPR010095 Transposase IS605, OrfB, C−terminal
7103 IPR008271 Serine/threonine−protein kinase, active site
728 IPR002893 Zinc finger, MYND−type
391 IPR000477 Reverse transcriptase domain
201 IPR006502 Protein of unknown function PDDEXK−like
1592 IPR009072 Histone−fold
3970 IPR036770 Ankyrin repeat−containing domain superfamily
909 IPR008979 Galactose−binding−like domain superfamily
5872 IPR032675 Leucine−rich repeat domain superfamily
597 IPR029787 Nucleotide cyclase
drm cel sac scp ptr tps esi aan atr osa ath ppa kni cme ccr olu o809 ota bpr mpu m299 cvu asp pse cnc64a hsp apr
cre gpe vca
Caulerpa taxifolia
Ulva mutabilis
significance
total count
Animals & Fungi
Heterokonts
Streptophyta
Rhodophyta
Prasinophytes
Trebouxiophyceae
Volvocales
−1 0 1 2 3 4 5
z-score
(normalized number
of genes with IPR)

Figure 6

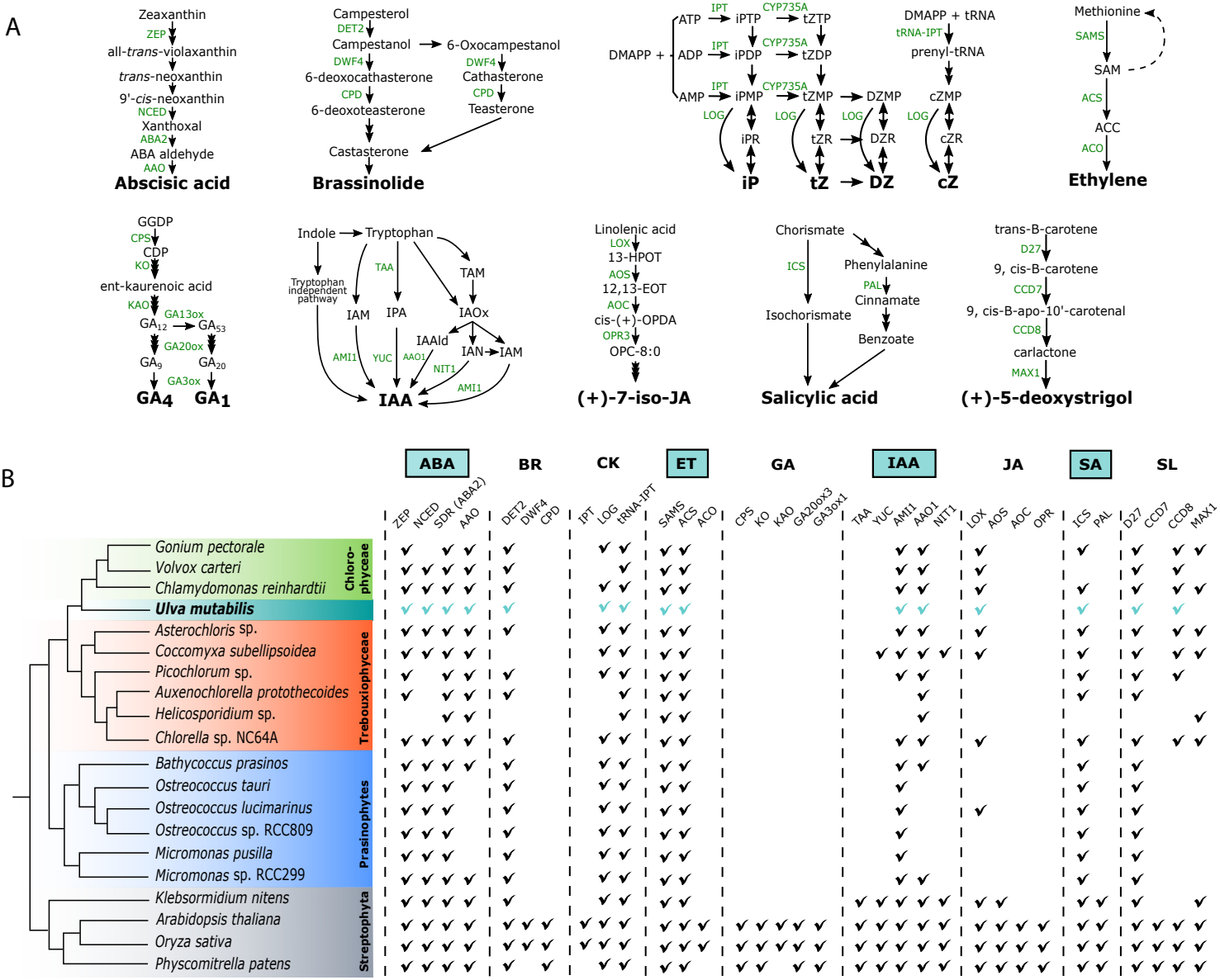

A
Zeaxanthin
ZEP
all-trans-violaxanthin
trans-neoxanthin
9'-cis-neoxanthin
NCED
Xanthoxal
ABA2
ABA aldehyde
AAO
Abscisic acid
Campesterol
DET2
Campestanol
6-Oxocampestanol
DWF4
6-deoxocathasterone
Cathasterone
CPD
6-deoxoteasterone
Teasterone
Castasterone
Brassinolide
DMAPP +
ATP
ADP
AMP
IPT
iPTP
iPDP
iPMP
CYP735A
tZTP
tZDP
tZMP
LOG
iPR
tZR
DZMP
DZR
iP
tZ
DZ
DMAPP + tRNA
tRNA-IPT
prenyl-tRNA
cZMP
cZR
cZ
Methionine
SAMS
SAM
ACS
ACC
ACO
Ethylene
GGDP
CPS
CDP
KO
ent-kaurenoic acid
KAO
GA13ox
GA12
GA53
GA20ox
GA9
GA20
GA3ox
GA4
GA1
Indole
Tryptophan
Tryptophan independent pathway
TAA
TAM
IAM
IPA
IAOx
IAAld
IAN
AMI1
YUC
AAO1
NIT1
IAA
Linolenic acid
LOX
13-HPOT
AOS
12,13-EOT
AOC
cis-(+)-OPDA
OPR3
OPC-8:0
(+)-7-iso-JA
Chorismate
ICS
Phenylalanine
PAL
Cinnamate
Isochorismate
Benzoate
Salicylic acid
trans-B-carotene
D27
9, cis-B-carotene
CCD7
9, cis-B-apo-10'-carotenal
CCD8
carlactone
MAX1
(+)-5-deoxystrigol
B
ABA
BR
CK
ET
GA
IAA
JA
SA
SL
ZEP
NCED
SDR (ABA2)
AAO
DET2
DWF4
CPD
IPT
LOG
tRNA-IPT
SAMS
ACS
ACO
CPS
KO
KAO
GA20ox3
GA3ox1
TAA
YUC
AMI1
AAO1
NIT1
LOX
AOS
AOC
OPR
ICS
PAL
D27
CCD7
CCD8
MAX1
Chloro-phyceae
Gonium pectorale
Volvox carteri
Chlamydomonas reinhardtii
Ulva mutabilis
Trebouxiophyceae
Asterochloris sp.
Coccomyxa subellipsoidea
Picochlorum sp.
Auxenochlorella protothecoides
Helicosporidium sp.
Chlorella sp. NC64A
Prasinophytes
Bathycoccus prasinos
Ostreococcus tauri
Ostreococcus lucimarinus
Ostreococcus sp. RCC809
Micromonas pusilla
Micromonas sp. RCC299
Streptophyta
Klebsormidium nitens
Arabidopsis thaliana
Oryza sativa
Physcomitrella patens

Figure 7

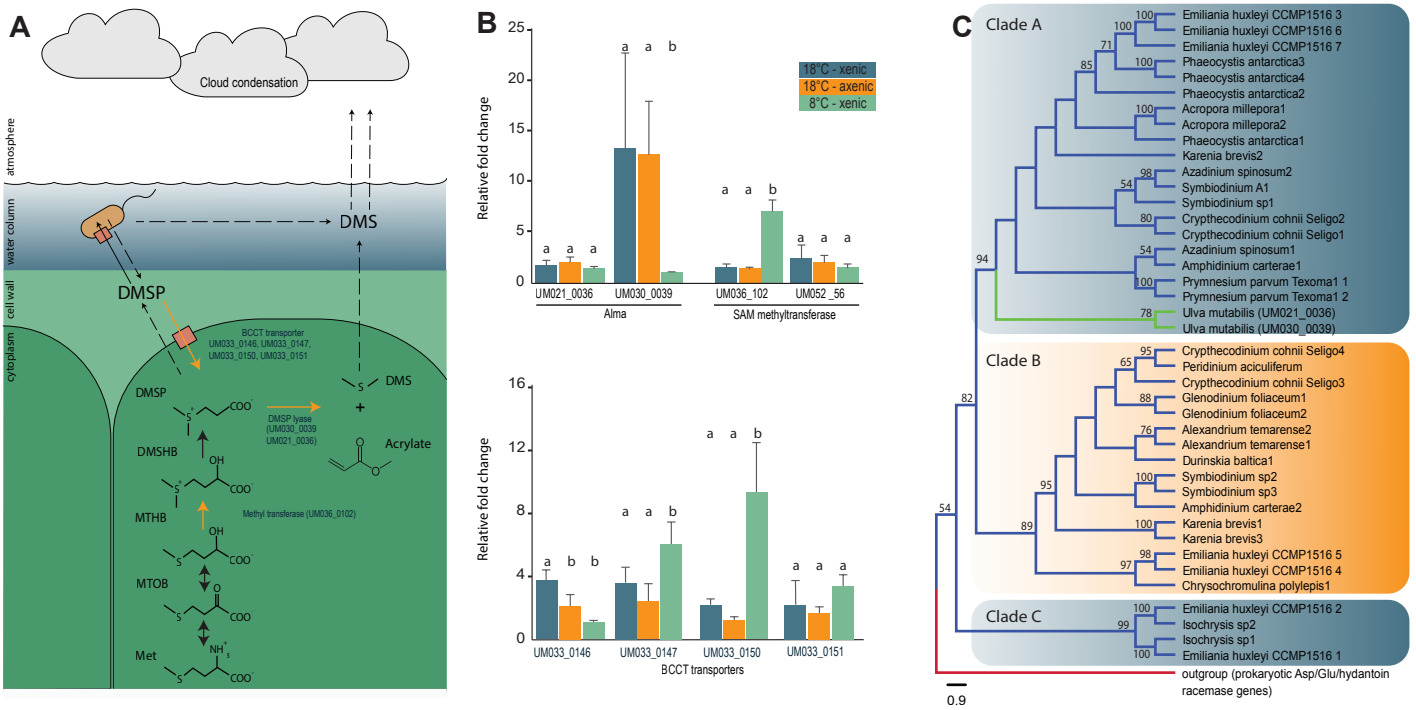

A
Cloud condensation
atmosphere
water column
cell wall
cytoplasm
DMS
DMSP
BCCT transporter
UM033_0146, UM033_0147,
UM033_0150, UM033_0151
DMSP
DMS
DMSP lyase
(UM030_0039
UM021_0036)
Acrylate
DMSHB
MTHB
Methyl transferase (UM036_0102)
MTOB
Met
B
Relative fold change
18°C - xenic
18°C - axenic
8°C - xenic
UM021_0036
UM030_0039
UM036_102
UM052_56
Alma
SAM methyltransferase
UM033_0146
UM033_0147
UM033_0150
UM033_0151
BCCT transporters
C
Clade A
Clade B
Clade C
Emiliania huxleyi CCMP1516 3
Emiliania huxleyi CCMP1516 6
Emiliania huxleyi CCMP1516 7
Phaeocystis antarctica3
Phaeocystis antarctica4
Phaeocystis antarctica2
Acropora millepora1
Acropora millepora2
Phaeocystis antarctica1
Karenia brevis2
Azadinium spinosum2
Symbiodinium A1
Symbiodinium sp1
Cryptechocodinium cohnii Seligo2
Cryptechocodinium cohnii Seligo1
Azadinium spinosum1
Amphidinium carterae1
Prymnesium parvum Texoma1 1
Prymnesium parvum Texoma1 2
Ulva mutabilis (UM021_0036)
Ulva mutabilis (UM030_0039)
Cryptechocodinium cohnii Seligo4
Peridinium aciculiferum
Cryptechocodinium cohnii Seligo3
Glenodinium foliaceum1
Glenodinium foliaceum2
Alexandrium temarense2
Alexandrium temarense1
Durinskia baltica1
Symbiodinium sp2
Symbiodinium sp3
Amphidinium carterae2
Karenia brevis1
Karenia brevis3
Emiliania huxleyi CCMP1516 5
Emiliania huxleyi CCMP1516 4
Chrysochromulina polylepis1
Emiliania huxleyi CCMP1516 2
Isochrysis sp2
Isochrysis sp1
Emiliania huxleyi CCMP1516 1
outgroup (prokaryotic Asp/Glu/hydantoin
racemase genes)
0.9

Figure S1. Phylogenetic position of the sequenced *Ulva mutabilis* strain. Related to Figure 1.

Maximum likelihood tree (log likelihood: -4386.98) of the *rbc*L gene (1,425 bp), based on a GTRGAMMA model with 25 rate categories and a single partition using RAxML v.8.2.8. Branch support results from a bootstrap analysis with 500 replicates. *Umbraulva olivascens* and *Ulvaria obscura* were used as outgroup.

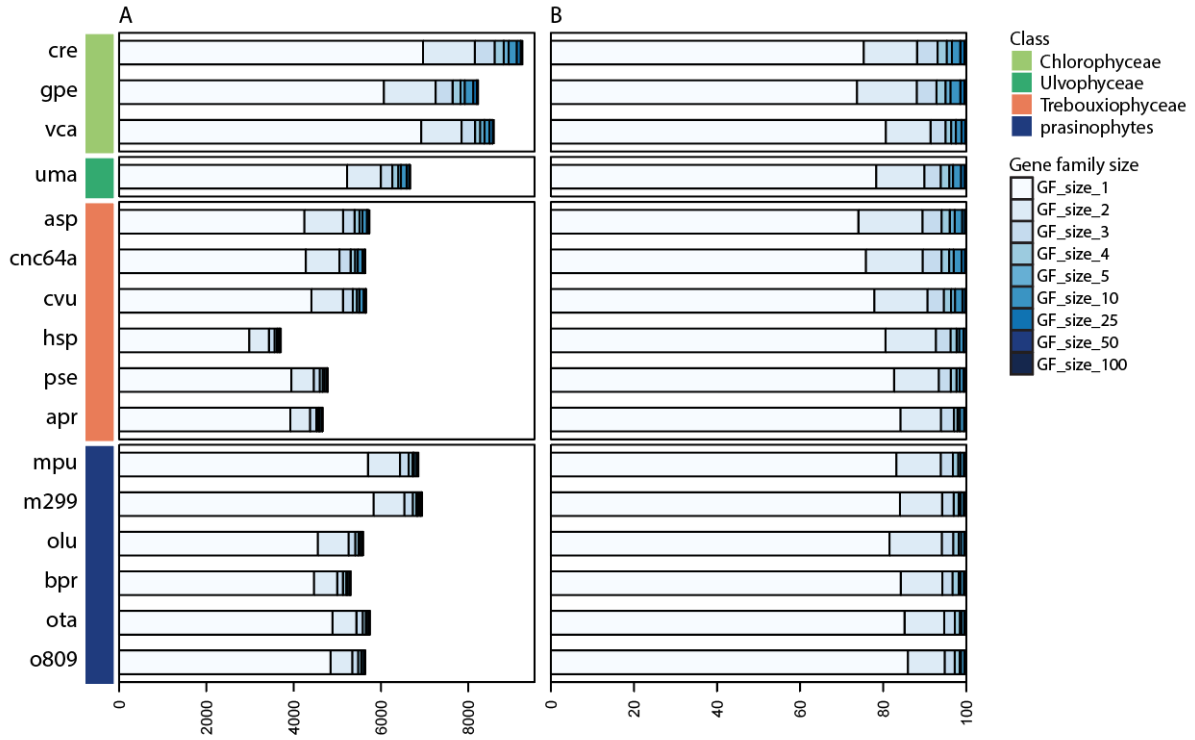


Figure S2. Distribution of gene family sizes in the green algal classes. Related to Figure 2.

A. actual numbers. B. percentages of total number of gene families. Abbreviations, see Table S8.

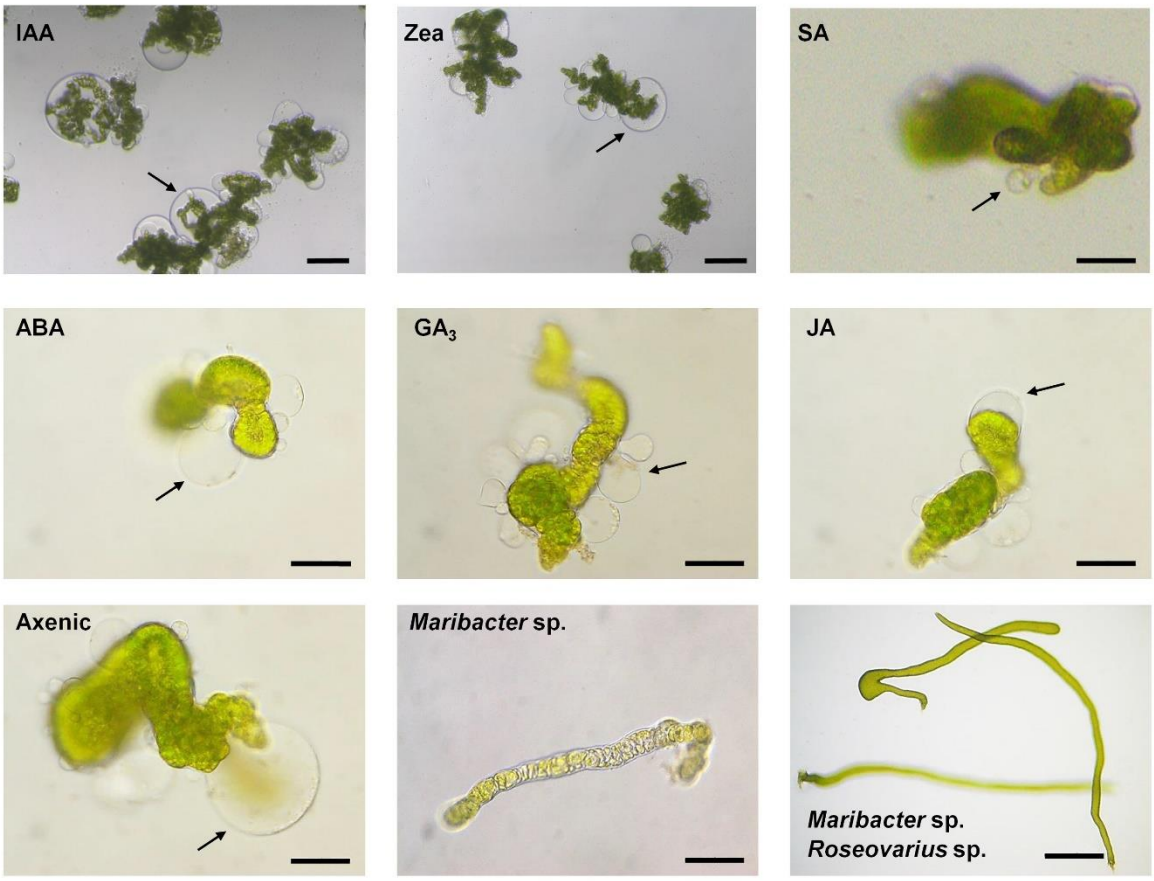


Figure S3. Morphogenetic bioassays of gametes inoculated with phytohormones. Related to Figure 6.

Representative algae of 2-weeks old germlings are shown. Protrusions of malformed cell walls are visible. Axenic gametes were inoculated with phytohormones (final concentration 10^{-9} mol/l). Phenotypes did not change at higher concentrations (10^{-6} mol/l). Top line: phytohormones IAA, Zea, SA, bars = 100 µm, middle line: phytohormones ABA, GA_3, JA, bars = 50 µm, bottom line: controls are shown under axenic conditions and inoculated with *Maribacter* sp. MS6 (bars = 50 µm). Growth and development of axenic cultures recovered upon inoculation with *Maribacter* sp. and *Roseovarius* sp. (bar = 500 µm). The combined application of all tested phytohormones did not recover the morphogenesis of *Ulva*.

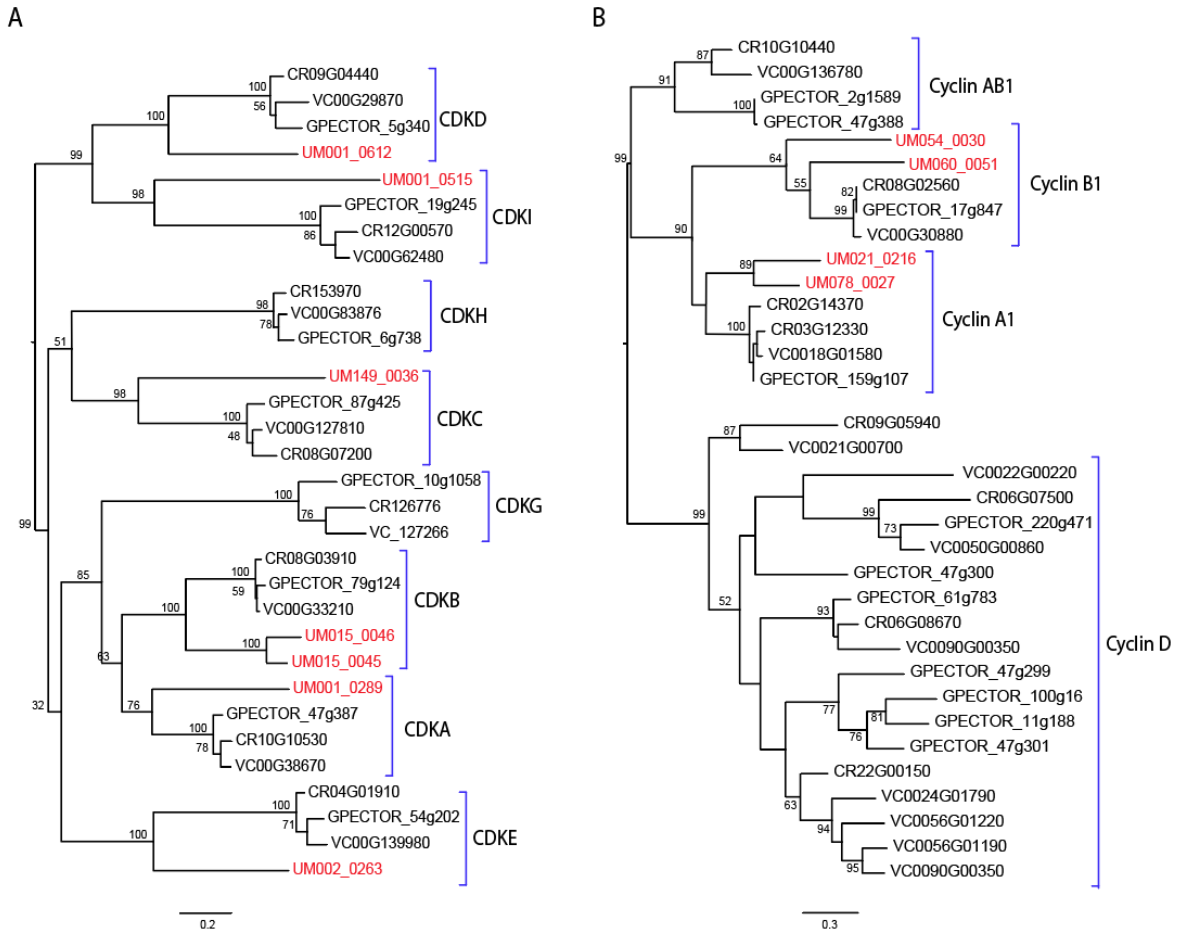
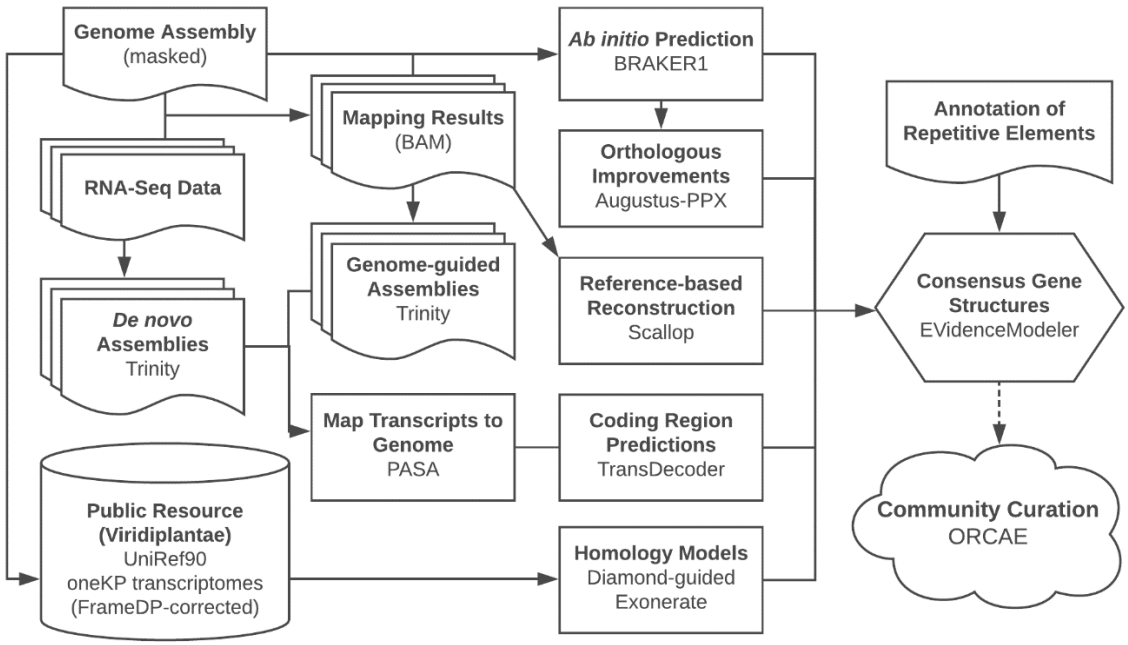


Figure S4. Phylogenetic analyses of cell cycle related genes of *Ulva mutabilis* and volvocine algae. Related to Figures 4 and 5.

A. Cyclin Dependent Kinases (CDKs); B. Cyclins. Amino acid alignments were analysed using an automatically selected LG+GAMMA model under the Maximum Likelihood optimality criterion in RAxML. Bootstrap values result from fast bootstrapping algorithm implemented in RAxML (v8.2.8). Assignment to different classes of CDK and cylin genes follows Hanschen et al. (2016). Abbreviations: *Chlamydomonas reinhardtii*, CR; Gonium pectoral, GPECTOR; *Volvox carteri*, VC.

A



B

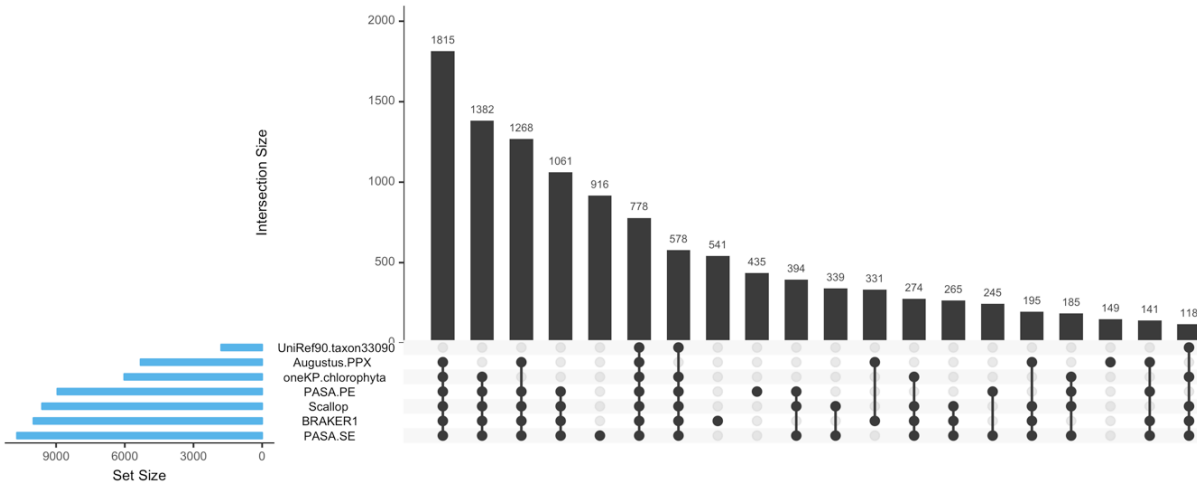


Figure S5. Gene prediction workflow for the *Ulva mutabilis* genome. Related to STAR Methods.

A. Workflow of the gene prediction. B. Contributions of the different approaches to the final gene models. PASA.PE: predicted coding regions of PASA alignments using assembled RNA-Seq data from paired-end stranded data; PASA.SE: predicted coding regions of PASA alignments using assembled RNA-Seq data from single-end stranded data. The top 20 intersects are shown.

Assembly metrics	Scaffolds	Contigs
Number of sequences (>= 1k bp)	318	381
Number of sequences (>= 5k bp)	315	378
Number of sequences (>= 10k bp)	312	372
Number of sequences (>= 25k bp)	276	322
Number of sequences (>= 50k bp)	243	274
Total length (>= 1k bp)	98,484,689	98,388,870
Total length (>= 5k bp)	98,477,802	98,381,983
Total length (>= 10k bp)	98,460,404	98,340,275
Total length (>= 25k bp)	97,848,084	97,481,650
Total length (>= 50k bp)	96,687,287	95,752,299
Largest sequence	3,623,364	2,868,306
Total length	98,484,689	98,388,870
GC (%)	57.2	57.2
N50	600,008	527,412
NG50	591,658	514,066
N75	340,001	289,826
NG75	330,723	267,228
L50	46	54
LG50	47	56
L75	100	117
LG75	103	122
Number of N's per 100 kbp	98	0

Table S1. Assembly metrics of the *Ulva* genome. Related to Figure 2.

Species	**Number of protein sequences**	**BUSCO**				**Core gene families**	
		Single-copy Complete	Duplicated	Fragmented	Missing	Number	GFscore
C. reinhardtii	17,741	285	6	7	5	1,812	0.999
C. subellipsoidea C169	9,629	265	5	14	19	1,787	0.991
D. salina	16,697	232	3	51	17	1,754	0.978
G. pectorale	16,290	243	4	39	17	1,779	0.991
M. pusilla ccmp1545	10,660	259	6	11	27	1,810	0.999
Micromonas RCC299	10,103	267	5	10	21	1,815	1.000
O. lucimarinus	7,796	230	26	11	36	1,815	1.000
U. mutabilis	12,924	264	2	14	23	1,717	0.968
V. carteri	14,247	281	6	6	10	1,795	0.994

Table S2. Gene space completeness revealed by eukaryotic BUSCO and Core Gene Family analysis. Related to Figure 2.

Class	Family	Total bases	% masked
LTR	Copia	13,323,328	13.53
	Gypsy	1,607,922	1.63
	Pao	141,541	0.14
	LTR	102,691	0.10
	DIRS	72,745	0.07
	Ngaro	49,613	0.05
	ERVL	46,795	0.05
LINE	LINE	3,433,947	3.49
	CRE-II	2,753,584	2.80
	CRE-Cnl1	1,408,967	1.43
	R1	720,404	0.73
	I	483,199	0.49
	Penelope	269,147	0.27
	L1-Tx1	223,981	0.23
	RTE-X	75,371	0.08
	CR1	35,377	0.04
DNA	hAT-Ac	39,812	0.04
	Dada	30,836	0.03
	DNA	29,444	0.03
	hAT-Tag1	25,903	0.03
	hAT-Charlie	25,772	0.03
	CMC-EnSpm	7,698	0.01
SINE	Alu	90,465	0.09
RC	Helitron	46,685	0.05
Reported	RS280 (EU256376.1)	575,835	0.58
	RS360 (EU263359.1)	123,676	0.13
Unknown	Unknown	8,994,613	9.13

Table S3. Summary of repetitive elements annotated in *Ulva mutabilis*. Related to Figure 2.

	LC-MS and GC/MS analyses (ng/g dry weight[1] or nL/h.g fresh weight[2])			Pathways		Produced by bacteria MS2/MS6
	Axenic (slender)	**Xenic (slender)**	**Xenic (wildtype)**	**Bio-synthesis**	**Intra-cellular signaling**	
Abscisic acid (ABA) [1]	2.03 ± 0.18	2.57 ± 0.24	5.33 ± 0.27	Complete	No	No/No
Auxin (IAA)[1]	48.0 ± 11.0	51.04 ± 5.79	58.38 ± 22.6	Complete	No	Yes/Yes
Ethylene (ET) [2]	0.35 ± 0.22	0.3 ± 0.20	ND	Complete	No	ND
Gibberellin (GA_3)[1]	< LOQ (traces)	< LOQ (traces)	No	Incomplete	No	Yes/Yes
Salicylic acid (SA)[1]	124.1 ± 14.45	28.68 ± 4.97	41.67 ± 2.24	Complete	No	Yes/Yes
Brassino-steroids	No	No	No	No	No	ND
Cytokinins (Zeatin, iP)	No	No	No	Incomplete	No	Yes/Yes (iP)
Jasmonate (JA)	No	No	No	No	No	ND
Strigolactones	No	No	No	No	No	ND

Table S4. Overview of phytohormones in axenic and non-axenic cultures. Related to Figure 6.

Xenic cultures of *Ulva mutabilis* (*wildtype* and *slender* strain) present a tripartite community with the bacterial strains *Roseovarius* sp. MS2 and *Maribacter* sp. MS6) and the presence of proteins involved in their biosynthesis or signaling. ND = not determined, LOQ = limit of quantification.

No.	Gene models	Annotation	Functional Category	Notes	Exon (nr)	GC (%)	Transcription (FPKM value)
1	UM005_0173	deacylase YbaK-like	Translation	Ensures translation fidelity	8	67.28	50.1
2	UM107_0008	50S ribosomal protein L18	Translation	Ribosome structural component. Mediates the attachment of the 5S RNA into the large ribosomal subunit in bacteria.	5	62.11	98.1
3	UM033_0031	ATP-dependent DNA helicase PcrA	DNA repair	DNA unwinding in bacteria. Belongs to UvrD/Rep helicase family.	8	59.42	73.4
	UM033_0032 (tandem dup.)				2	57.74	18.4
4	UM042_0001	DNA-3-methyladenine glycosylase	DNA repair	DNA alkylation repair	6	56.62	
	UM084_0052				5	52.05	24.6
5	UM013_0142	glucuronyl hydrolase	Carbohydrate metabolism	Glycosyl Hydrolase Family 88; Unsaturated glucuronyl hydrolase catalyzes the hydrolytic release of unsaturated glucuronic acids from oligosaccharides (EC:3.2.1.-) produced by the reactions of polysaccharide lyases.	7	64.01	17.7
	UM034_0033				5	66.36	32.5
6	UM035_0080	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	Cell wall metabolism	MurF. Involved in cell wall formation. Catalyzes the final step in the synthesis of UDP-N-acetylmuramoyl-pentapeptide, the precursor of murein.	3	53.31	18.2
7	UM016_0037	Haem peroxidase	Response to stress	Reduces hydrogen peroxide	8	61.45	31.1
	UM001_0001				4	62.89	0.5
	UM001_0017				4	60.94	0
	UM001_0023				5	61.36	1.9
	UM001_0025				5	60.99	36.7
	UM001_0027				3	54.85	0.2
	UM001_0028				3	61.55	7

	UM001_0030				1	61.43	0
	UM001_0033				4	61.38	0
	UM017_0003				5	61.21	0.3
	UM017_0004				5	61.19	0
	UM017_0005				5	62.63	0
	UM018_0064				5	63.54	4.3
	UM018_0067				7	67.09	210.6
	UM024_0048				4	62.83	3.1
	UM024_0060				6	59.73	32.6
	UM049_0004				5	60.16	0.1
	UM049_0029				6	64.81	56.5
	UM061_0028				8	60.06	86.4
	UM077_0050				5	64.65	1
	UM104_0025				4	62.96	26.4
	UM104_0027				5	59.9	1.9
	UM104_0028				4	62.48	0.5
	UM104_0029				3	62.19	1
	UM104_0030				2	63.98	0.0
	UM154_0005				4	61.3	0.9
	UM160_0003				6	60.89	0.8
	UM169_0003				7	63.88	1.1
	UM180_0001				4	59.59	0.1
	UM180_0002				2	58.5	2.4
	UM180_0003				4	61.2	1.3
	UM180_0006				5	62.04	0.5
	UM180_0007				2	60.7	0.0
	UM199_0002				2	62.17	0.0
	UM290_0001				4	62.06	0.0
	UM290_0002				3	60.33	0.0

8	UM109_0045	cupin	other	The cupin superfamily of proteins is functionally extremely diverse. It was named on the basis of the conserved β-barrel fold	2	58.05	32.9
9	UM052_0037	D-lactate dehydrogenase	other	Catalyzes interconversion of pyruvate and lactate	4	61.19	1.8
10	UM031_0093	serine/threonine protein kinase	other	catalytic domain	7	66.53	47.5
	UM031_0094 (tandem dup.)				3	65.73	30.7
	UM031_0095 (tandem dup.)				6	67.22	13.3
	UM031_0097 (tandem dup.)				6	67.6	3.4
11	UM023_0015	multifunctional 2',3'-cyclic-nucleotide 2'-phosphodiesterase/5'-nucleotidase/3'-nucleotidase	other	nucleotide scavenging and transport	9	66.48	27.1
	UM035_0119				9	57.34	22.3
	UM042_0110				3	68.56	0.0
	UM042_0111				10	65.05	11.4
12	UM094_0024	serine/threonine protein kinase	other	catalytic domain	3	68.55	1.2
	UM001_0092				3	67.5	24.2
	UM001_0100				3	67.65	42.1
	UM006_0116				4	68.69	6.7
	UM011_0009				3	59.68	5.2
	UM025_0014				3	66.59	26.2
	UM173_0008				2	67.24	14.2
13	UM009_0111	DUF3179 domain-containing protein	unknown function		1	55.74	26

Table S5. Horizontal gene transfer candidates in *Ulva mutabilis.* Related to "Macroalgal-bacterial interactions" section.

Sequencing Platform	Source and Type	Strain	Read Number	Total bases (Gbp)
PacBio RSII (P6 chemistry)	DNA (>8 kb)	WT	770,598	6.91
Illumina MiSeq (2x250 bp paired-end)	DNA (insert size 350, 550 bp)	WT	16,398,654	2.92
Illumina HiSeq 2500 (2x125 bp paired-end)	DNA (insert size 2kb)	WT	170,149,048	17.75
Illumina HiSeq 2500 (2x125 bp paired-end)	DNA (insert size 5kb)	WT	166,928,626	17.50
Illumina HiSeq 2500 (2x125 bp paired-end)	DNA (insert size 10kb)	WT	126,821,228	13.13
Illumina HiSeq 2500 (2x125 bp paired-end)	RNA (FR-second strand)	WT, Sl	397,961,040	44.67
NextSeq 550 (150 bp single-end)	RNA (reverse strand)	Sl	164,378,893	22.56

Table S6. DNA and RNA sequencing libraries. Related to STAR Methods.
WT: wild-type strain; SL: slender strain

Species	Source	PubmedID	Abbrev.
Aureococcus anophagefferens	JGI 1.0	21368207	aan
Asterochloris sp. *Cgr/DA1pho v2.0*	JGI 7.45.13	/	asp
Auxenochlorella protothecoides	Beijing Genomics Institute 1.0	25012212	apr
Arabidopsis thaliana	TAIR10	11130711	ath
Amborella trichopoda	Amborella v1	24357323	atr
Bathycoccus prasinos	Ghent University	22925495	bpr
Chondrus crispus	ENSEMBL release 28	23536846	ccr
Caenorhabditis elegans	ENSEMBL release 81	9851916	cel
Cyanidioschyzon merolae	Tokyo University	15071595	cme
Chlorella sp *NC64A*	JGI 1.0	20852019	cnc64a
Chlamydomonas reinhardtii	JGI 5.5 (Phytozome 10.2)	17932292	cre
Coccomyxa subellipsoidea C-169	JGI 2.0 (Phytozome 10.2)	22630137	cvu
Drosophila melanogaster	ENSEMBL release 81	10731132	drm
Ectocarpus siliculosus	Ghent University	20520714	esi
Gonium pectorale	GenBank (LSYV01000000)	27102219	gpe
Homo sapiens	ENSEMBL release 81	11181995	hom
Helicosporidium sp.	Illinois University 1.0	24809511	hsp
Klebsormidium nitens	Tokyo Inst. Technology	24865297	kni
Micromonas pusilla strain CCMP1545	Ghent University	24273312	mpu
Micromonas sp. *RCC299*	JGI 3.0	19359590	m299
Ostreococcus lucimarinus	JGI 2.0	17460045	olu
Ostreococcus sp. *RCC809*	JGI 2.0	/	o809
Ostreococcus tauri	Ghent University v2.0	25494611	ota
Oryza sativa	MSU RGAP 7	16100779	osa
Physcomitrella patens	Phytozome 9.1 (v1.6)	18079367	ppa
Picochlorum sp. *SENEW3 (SE3)*	Rutgers University 1.0	24965277	pse
Phaeodactylum tricornutum	ASM15095v2	18923393	ptr
Saccharomyces cerevisiae strain S288C	ENSEMBL release 81	8849441	sac
Schizosaccharomyces pombe	ENSEMBL fungi release 28	11859360	scp
Thalassiosira pseudonana	JGI 3.0	15459382	tps
Ulva mutabilis	/	/	uma
Volvox carteri	JGI 2.0 (Phtyozome 10.2)	20616280	vca

Table S7. Species list and genome versions used for annotation and comparative genomics in a custom PLAZA platform. Related to Figures 4, 5 and STAR Methods.

name	target gene identifier	primer sequence (5'->3')
p111f	UM008_0183 (Ubiquitin)	CCCTCGAAGTGGAGTCTTCTGAC
p112r	UM008_0183 (Ubiquitin)	AAGTGTGCGGCCATCCTCTA
p173f	UM010_0003 (PP2A 65 kDa regulatory subunit A)	GGCAACTGCAGGAGCAATTCT
p160r	UM010_0003 (PP2A 65 kDa regulatory subunit A)	CCTCAGAAGCAACCTCGACCAT
p169f	UM030_0039 (DMSP lyase)	TTCGACGACAAAGAGAAGATCGCA
p170r	UM030_0039 (DMSP lyase)	TAGCGGTCCTTCTCCAGGTC
p167f	UM021_0036 (DMSP lyase)	CAAGCCTGTGCTGCTCTCG
p171r	UM021_0036 (DMSP lyase)	GTGGCCGTTTGCCGTAAAGA
p174f	UM033_0146 (BCCT transporter)	CAACGCCAGGCTCCAAGAC
p175r	UM033_0146 (BCCT transporter)	CACCCATGCTTGCTGTGTGA
p180f	UM033_0147 (BCCT transporter)	GGGCGGTCAACATGTCGTT
p181r	UM033_0147 (BCCT transporter)	GGACTTCATCGTCATCGAGAACC
p184f	UM033_0150 (BCCT transporter)	AGGCCATCAATCTGACCGGATT
p185r	UM033_0150 (BCCT transporter)	AGACTTGAGTGTCATGGCAAAGC
p186f	UM033_0151 (BCCT transporter)	GGCCAGAGAAGTCAAATGCAGAG
p187r	UM033_0151 (BCCT transporter)	ACCACAAAGATGTTCTGAAGGCC
p190f	UM036_0102 (methyltransferase)	GCAATGGCTGAAGCATGCG
p191r	UM036_0102 (methyltransferase)	GCCCTTGCCTGAGCCTAAATC
p192f	UM052_0056 (methyltransferase)	ATCAACTTTGGCAACTCCCTGC
p193r	UM052_0056 (methyltransferase)	GTTGGGGCCAAGCAGAATG

Table S8. Target genes and associated primers for qPCR analysis of DMSP-related genes. Related to Figure 7.

Excel table - annotation

Click here to access/download

Supplemental Videos and Spreadsheets

Data S1.xlsx